

GenCore version 5.1.4 ps 4578
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2003, 15:54:33 ; Search time 293 Seconds
(without alignments)
4904.289 Million cell updates/sec

Title: US-09-315-355a-47

Perfect score: 1853
Sequence: 1 gatccctgagcgtggtgcag.....aaagtaattccttaaacat 1853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	150	8.1	150	10 US-09-880-107-2725	Sequence 2725, Ap
C 2	53.2	2.9	2895	10 US-09-938-842A-2540	Sequence 2540, Ap
C 3	52.8	2.8	766	10 US-09-864-761-19608	Sequence 19608, A
C 4	52.8	2.8	1944	10 US-09-864-761-2825	Sequence 2825, Ap
C 5	49	2.6	1959	10 US-09-864-761-4012	Sequence 4012, Ap
C 6	47.8	2.6	583	10 US-09-864-761-20772	Sequence 20772, A
C 7	47	2.5	1626	10 US-09-764-848-19	Sequence 19, Appl
C 8	46.2	2.5	58985	9 US-09-901-152-3	Sequence 3, Appl
C 9	44.6	2.4	4439	10 US-09-864-761-20174	Sequence 20174, A
C 10	44.2	2.4	2142	10 US-09-917-800A-1567	Sequence 1567, Ap
C 11	43.8	2.4	888	10 US-09-770-445-502	Sequence 502, App
C 12	43.6	2.4	327	10 US-09-864-761-28059	Sequence 28059, A
C 13	43.6	2.4	575	10 US-09-864-761-20733	Sequence 20733, A
C 14	43.6	2.4	1969	10 US-09-864-761-3972	Sequence 3972, Ap
C 15	43.4	2.3	720	9 US-10-101-487-74	Sequence 74, Appl
C 16	43.4	2.3	720	9 US-10-101-487-76	Sequence 887, Appl
C 17	43	2.3	8895	10 US-09-764-853-887	Sequence 937, App
C 18	43	2.3	8895	10 US-09-764-853-937	Sequence 886, App
C 19	43	2.3	9656	10 US-09-764-853-886	

C 20	43	2.3	9656	10 US-09-764-853-933	Sequence 933, App
C 21	42.8	2.3	522	9 US-10-101-487-71	Sequence 71, Appl
C 22	42.8	2.3	530	9 US-10-101-487-73	Sequence 73, Appl
C 23	42.8	2.3	530	9 US-10-101-487-69	Sequence 69, Appl
C 24	42.8	2.3	554	9 US-10-101-487-106	Sequence 106, Appl
C 25	42.4	2.3	565	10 US-09-864-761-6633	Sequence 6633, Ap
C 26	42.2	2.3	1920	10 US-09-906-393A-35	Sequence 35, Appl
C 27	42.2	2.3	1958	10 US-09-791-406-3	Sequence 3, Appl
C 28	42.2	2.3	110096	10 US-09-880-107-1542	Sequence 1542, Ap
C 29	41.4	2.2	460	10 US-09-864-761-2587	Sequence 2587, Ap
C 30	41.4	2.2	466	10 US-09-864-761-98	Sequence 8168, Appl
C 31	41.4	2.2	582	10 US-09-864-761-8188	Sequence 1361, Ap
C 32	41.2	2.2	475	10 US-09-864-761-1361	Sequence 18121, A
C 33	41.2	2.2	512	10 US-09-864-761-18121	Sequence 18121, A
C 34	40.4	2.2	345	10 US-09-822-263-7	Sequence 332, App
C 35	40.2	2.2	520	9 US-10-184-644-332	Sequence 20742, A
C 36	40.2	2.2	632	10 US-09-864-761-20742	Sequence 3981, Ap
C 37	40.2	2.2	1983	10 US-09-864-761-3981	Sequence 22453, A
C 38	39.6	2.1	422	10 US-09-864-761-22453	Sequence 5690, Ap
C 39	39.6	2.1	484	10 US-09-864-761-5690	Sequence 19488, A
C 40	39.2	2.1	660	10 US-09-864-761-19488	Sequence 2772, Ap
C 41	39.2	2.1	962	10 US-09-864-761-2772	Sequence 2650, Ap
C 42	39.2	2.1	1881	9 US-09-938-842A-2650	Sequence 9004, Ap
C 43	39	2.1	501	10 US-09-864-761-9004	Sequence 11, Appl
C 44	38.8	2.1	342	10 US-09-822-263-11	Sequence 3, Appl
C 45	38.8	2.1	555	10 US-09-822-263-3	

ALIGNMENTS

RESULT 1	US-09-880-107-2725/C
Sequence 2725, Application US/09880107	
Patent No. US20020142981A1	
GENERAL INFORMATION:	
APPLICANT: Horne, Darci T.	
APPLICANT: Vockley, Joseph G.	
APPLICANT: Scherf, Uwe	
APPLICANT: Gene Logic, Inc.	
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer	
FILE REFERENCE: 44921-5028-WO	
CURRENT APPLICATION NUMBER: US/09/880,107	
CURRENT FILING DATE: 2001-06-14	
PRIOR APPLICATION NUMBER: US 60/211,379	
PRIOR FILING DATE: 2000-06-14	
PRIOR APPLICATION NUMBER: US 60/237,054	
PRIOR FILING DATE: 2000-10-02	
NUMBER OF SEQ ID NOS: 3950	
SOFTWARE: PatentIn Ver. 2.1	
SEQ ID NO 2725	
LENGTH: 150	
TYPE: DNA	
ORGANISM: Homo sapiens	
FEATURE:	
OTHER INFORMATION: Genbank Accession No. US20020142981A1 N92948	
US-09-880-107-2725	
Query Match	8.1%; Score 150; DB 10; Length 150;
Best Local Similarity	100.0%; Pred. No. 6.6e-33;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 660 GGAATGCTGAATTTGATCTGACCCAGATGATTCTACTGGAATTAATGATGCTGTAGG 719	
DB 150 GGAATGCTGAATTTGATCTGACCCAGATGATTCTACTGGAATTAATGATGCTGTAGG 91	
QY 720 AAACATGACCCCTGTTATGAGTGTGGACCTTGATATAGTGACCTTTAGACCCAGT 779	
DB 90 AAACATGACCCCTGTTATGAGTGTGGACCTTGATATAGTGACCTTTAGACCCAGT 31	
QY 780 CTTACACCTCGAAGTAACTTTCAAAAA 809	
DB 30 CTTACACCTCGAAGTAACTTTCAAAAA 1	

QY 500 ATGAACGGAGATTTCCTTGATTAAAGCCCACTGATTAATCTTATAGTTGGCCGACTG 558
 Db 145 ATGAAGAGAGAGTTGCCATGCTGATGATGCTGATTAAGCCGATGTAAGTGTATG 86
 QY 560 AACAGACCCAGTGCATTTAGAGTGCATGTTTAATCAAGAAGAAGACTCTTTATG 619
 Db 85 ATGATTAACGAGGAAGAGATGCTATGATGATGATGATGAGAGAGTTGGTATGATG 26
 QY 620 TACACCATGATA 631
 Db 25 ATGTGATGATA 14

```

1 RESULT 4
2 US-09-864-761-2825/c
3 : Sequence 2825, Application US/09864761
4 Patent No. US20020048763A1
5 GENERAL INFORMATION:
6 APPLICANT: Penn, Sharon G.
7 APPLICANT: Rank, David R.
8 APPLICANT: Hanzel, David K.
9 APPLICANT: Chen, Wensheng
10 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
11 FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
12 FILE REFERENCE: Aecmica-X-1
13 CURRENT APPLICATION NUMBER: US/09/864,761
14 CURRENT FILING DATE: 2001-05-23
15 PRIOR APPLICATION NUMBER: US 60/180,312
16 PRIOR FILING DATE: 2000-02-04
17 PRIOR APPLICATION NUMBER: US 60/207,456
18 PRIOR FILING DATE: 2000-05-26
19 PRIOR APPLICATION NUMBER: US 09/632,366
20 PRIOR FILING DATE: 2000-08-03
21 PRIOR APPLICATION NUMBER: GB 24263.6
22 PRIOR FILING DATE: 2000-10-04
23 PRIOR APPLICATION NUMBER: US 60/236,359
24 PRIOR FILING DATE: 2000-09-27
25 PRIOR APPLICATION NUMBER: PCT/US01/00666
26 PRIOR FILING DATE: 2001-01-30
27 PRIOR APPLICATION NUMBER: PCT/US01/00667
28 PRIOR FILING DATE: 2001-01-30
29 PRIOR APPLICATION NUMBER: PCT/US01/00664
30 PRIOR FILING DATE: 2001-01-30
31 PRIOR APPLICATION NUMBER: PCT/US01/00669
32 PRIOR FILING DATE: 2001-01-30
33 PRIOR APPLICATION NUMBER: PCT/US01/00665
34 PRIOR FILING DATE: 2001-01-30
35 PRIOR APPLICATION NUMBER: PCT/US01/00668
36 PRIOR FILING DATE: 2001-01-30
37 PRIOR APPLICATION NUMBER: PCT/US01/00663
38 PRIOR FILING DATE: 2001-01-30
39 PRIOR APPLICATION NUMBER: PCT/US01/00662
40 PRIOR FILING DATE: 2001-01-30
41 PRIOR APPLICATION NUMBER: PCT/US01/00661
42 PRIOR FILING DATE: 2001-01-30
43 PRIOR APPLICATION NUMBER: PCT/US01/00670
44 PRIOR FILING DATE: 2001-01-30
45 PRIOR APPLICATION NUMBER: US 60/234,687
46 PRIOR FILING DATE: 2000-09-21
47 PRIOR APPLICATION NUMBER: US 09/608,408
48 PRIOR FILING DATE: 2000-06-30
49 PRIOR APPLICATION NUMBER: US 09/774,203
50 PRIOR FILING DATE: 2001-01-29
51 NUMBER OF SEQ ID NOS: 49117
52 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
53 SEQ ID NO 2825
54 LENGTH: 1944
55 TYPE: DNA
56 ORGANISM: Homo sapiens
57 FEATURE:
58 OTHER INFORMATION: MAP TO AL008720.1
59 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
60 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89

```

```

; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-2825

```

Query Match	2.8%	Score 52.8	DB 10	Length 1944
Best Local	Similarly	48.1%	Pred. No. 0.00016	
Matches 150	Conservative	0	Mismatches 162	Indels 0
				Gaps 0

Oy	320	ATGGGACCCAGAGATACAGACGCTTGATATATGATAGCTGGCTGAGACCACTTAG	379
Dd	552	ATGTGGCAATATGTAATGATGTGTGATAGAGAGAGGATGGCGATGATGGTATGATG	503
Oy	380	ATTAATATGATGAGGAAGTGAACCCAGATGCTGAGACTCTTGGTGAATCTCTTGGGTC	439
Dd	502	ATGATCTTGGCGGATGMAATATCAATATATGAAGAGAGCTTGGTGAATGGTATGATG	443
Oy	440	TTACGGTCTACGGAGATATGATCAAGATCCTTACGTTACTCTGAANAGATACAGACAAT	499
Dd	442	TTAATGGTATGTAATGATGTGTATGAAGATGATGATGTTGGCGATGAAGATGACATG	383
Oy	500	ATGAACGTGAAGATTTCTTGATTTAAGCCCACTGATTAATCTTATAGTTTGGCCGACTG	555
Dd	382	ATGAAGAGAGAGTTGGCGATGGTATGATGTGTGATTAATGGCGATGATAGCATGTATG	322
Oy	580	AACAGGACCACTGCATTTNAGTGCATGTTTATATCAAGAAAGACCTCTTTATG	619
Dd	322	ATGATTAACAGAGAAAGAGATGGTATGATGTGTATGATGAGAGAGTTGTGTATGATG	263
Oy	620	TACACCATGATA	631
Dd	262	ATGGTATGATAA	251

```

, RESULT 5
, US-09-864-761-4012
, Sequence 4012, Application US/09864761
, Patent No. US20020048763A1
, GENERAL INFORMATION:
, APPLICANT: Penn, Sharon G.
, APPLICANT: Rank, David R.
, APPLICANT: Hanzel, David K.
, APPLICANT: Chen, Wensheng
, TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
, TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
, FILE REFERENCE: Aecomica-X-1
, CURRENT APPLICATION NUMBER: US/09/864,761
, CURRENT FILING DATE: 2001-05-23
, PRIOR APPLICATION NUMBER: US 60/180,312
, PRIOR FILING DATE: 2000-02-04
, PRIOR APPLICATION NUMBER: US 60/207,456
, PRIOR FILING DATE: 2000-05-26
, PRIOR APPLICATION NUMBER: US 09/632,366
, PRIOR FILING DATE: 2000-08-03
, PRIOR APPLICATION NUMBER: GB 24263.6
, PRIOR FILING DATE: 2000-10-04
, PRIOR APPLICATION NUMBER: US 60/236,359
, PRIOR FILING DATE: 2000-09-27
, PRIOR APPLICATION NUMBER: PCT/US01/00666
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/00667
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/00664
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/00669
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/00665
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/00668
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/00663
, PRIOR FILING DATE: 2001-01-30

```

[illegible]

Db	125	GATGATGGTATGGTATGATGATGATGATGATGGGAAAGATGACAGTGATGGT	184
Qy	436	GGCTTACGGCTACGGAGTAATGATCCAAAGTACTCTGAAAGATACGAA	455
Db	185	GGTATGATGACGATGAGGATGAGGGGATGATACAGTATGATTTGGATGGTAT	244
Qy	496	CAATATGAACTGAAAGATTTCTTATTAAAGCCAGTATATCTTAATAGTTTGGCCGA	555
Db	245	GATGATGATGGTATGATGGGGATGATGATACAGTATATGATGGTATGATGAT	304
Qy	556	GCTGAACAGGACCGATGCAATTTAGAGTGCATTTTAATAACAAGAAAGACTCTTTT	615
Db	305	GATGGTATGACGGGGATGATGATGAGAGTGATGATGGTATGATGATGATGT	364
Qy	616	TATGTACACCATGAT	630
Db	365	GATGATGACGGTAT	379

```

RESULT 7
US-09-764-848-19
; Sequence 19, Application US/09764848
; Patent No. US2002007270A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1708
; CURRENT APPLICATION NUMBER: US/09/764,848
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-848-19

```

Query Match	2.5%;	Score 47;	DB 10;	Length 1626;
Best Local Similarity	51.7%;	Pred. No. 0.0066;		
Matches 104;	Conservative 1;	Mismatches 96;	Indels 0;	Gaps 0;

QY	208	GAGAAATTCGCAAGAAGGCTGGCAGTGATGTAAGAAGGAGACAGCAGCTTCAGAA	267
Db	372	GAAATATATAGAAAGAGAAATCTGATGCTGAATATATATGATGCTTGGACCTAAAGAAA	431
QY	268	GATGGCATGCAGAGTGCACGCACCCAGGACACGCCAAGAGAGCCCTCGAGAGATGTGC	327
Db	432	AAATATCAAGAGGGAAGGACGTTGGTCTCGCATCTATATTAAAGAAAGTTGAAGATTAAGAA	491
QY	328	CCAGAGGATGCAGAGACGCTTGATGATGATGATGAGCTGGCTGATGACCTTAGATAATAT	387
Db	492	TCAAGAGGAGAAAGAGAGATGAGAGATGAAGATCTTTCAATATTAATGATGATGAGAT	551
QY	388	GATGAGGAAGGTGACCCAGAT	408
Db	552	GAGATGAGATGACGCTGAT	572

```

RESULT 8
US-09-901-152-3/c
Sequence 3, Application US/09901152
Publication No. US20030022824A1
GENERAL INFORMATION:
APPLICANT: HU, Song et al.
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: CL001248
CURRENT APPLICATION NUMBER: US/09/901,152
CURRENT FILING DATE: 2001-07-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq For Windows Version 4.0

```

```

; SEQ ID NO 3
; LENGTH: 58985
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(58985)
; OTHER INFORMATION: n = A,T,C or G
US-09-901-152-3

```

Query Match	2.5%;	SCORE	46.2;	DB	9;	Length	58985;
Best Local Similarity	50.7%;	Pred. No.	0.11;				
Matches	11;	Conservative	0;	Mismatches	106;	Indels	0;
				Gaps			0;

[illegible]

```

RESULT 9
US-09-864-761-20174/c
Sequence 20174, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30

```

Prior Filing Date: 2001-05-11	
Prior Application Number: US 60/290,645	
Prior Filing Date: 2001-05-15	
Prior Application Number: US 60/292,336	
Prior Filing Date: 2001-05-22	
Prior Application Number: US 60/295,798	
Prior Filing Date: 2001-06-06	
Prior Application Number: US 60/297,457	
Prior Filing Date: 2001-06-13	
Prior Application Number: US 60/298,884	
Prior Filing Date: 2001-06-19	
Prior Application Number: US 60/303,459	
Prior Filing Date: 2001-07-09	
Number of SEQ ID NOS: 1740	
Software: PatentIn Ver. 2.1	
SEQ ID NO 1567	
Length: 2142	
Type: DNA	
Organism: Rattus norvegicus	
Feature:	
Other Information: Genbank Accession No. US20020119462A1 NM_012749	
US-09-917-800A-1567	
Query Match	2.4%; Score 44.2; DB 10;
Best Local Similarity	47.9%; Pred. No. 0.05;
Matches 127; Conservative	0; Mismatches 138; Indels 0; Gaps 0;

	Query Match	2.4%	Score 44.2	DB 10	Length 2142
	Best Local Similarity	47.9%	Pred. No. 0.05		
	Matches 127	Conservative 0	Mismatches 138	Indels 0	Gaps 0
QY	149	CACGAGCAAGTAGACGCTGATAAAGAAAGTAAACGCTCATTTGCTGAGCCAAAGC	208		
DB	551	CTCTCGCTTCAGAGAGATGAGATGAGAAAGATGATGATGATGATGATGATGATGATG	610		
DB	208	ACAAATTTCAACAGCAACCTTGTGCTGATGATGATGATGATGATGATGATGATGATG	268		

Oy	149	CACCAAGCAGAGTGACCTAGTAAAGAAAGTAAAGCCCTCATCTGCGAGGCAAGG	208
Dg	551	CTCTCGCTTCAGAGATGAGAGTATGAGAAAGTATGTATGAAGATGATGATGATG	610
Oy	209	AGAAATTGCAGAAGAGTGTTGGCAGTGTGAAGAGGAGACAGCAGTCTTCAGAAAG	268
Db	611	ATGAAGAGAGGAGAGAGAGATGATCTGTGAGAAAGAAATTATGCAATCACACAGCA	670
Oy	269	ATGCGATGCAGAGTGCACGCCACCCAGGACGCCCAAGAGACCCCTGAGAGATGTGACC	328
Db	671	AAGAAAGAAAACTCTGCAAAAGTTTCTCTGTGAAGGCCAAGTGTGGCCAGAGAGG	730
Oy	329	CAGAGATGACAGAGACCGTTGATGATGATGAGCTGGCTGATACGATTAATTAATG	388
Db	731	AGGAAGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	790
Oy	389	ATGAGGAAGGTGATCCAGATGCTGA	413
Db	791	ACGATGAAGATGAGAGATGAGGAAGA	815

RESULT 11
US-09-770-445-502
Sequence 502, Application US/09770445
Patent No. US2002023281A1
GENERAL INFORMATION:
APPLICANT: Gorlach, Jorn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kricker, Maria
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurbn, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis

TITLE OF INVENTION: thaliana
FILE REFERENCE: 2023US (PARA-012PRV)
CURRENT APPLICATION NUMBER: US/09/770,445
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/178,472
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 502
LENGTH: 888
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(888)
OTHER INFORMATION: n = A,T,C or G
US-09-770-445-502

Query Match 2.4%; Score 43.8; DB 10; Length 888;
Best Local Similarity 46.8%; Pred. No. 0.037;
Matches 138; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 131 GCGGCGTGGCCAAAGACACCCAGACAGGTAGAGCTGAGTAAAGAAAGTAAACGCC 190
DB 35 GCGACGACATAGTGTGAAGCAAGTTAAAGCGTAGGTGAAGATTAGCGACGAAG 94
QY 191 TCATTGCTGAGCAAGAGAAATTGCAAGAGAGGTGGTGGCAGTATGAGAGAGAGA 250
DB 95 ACAGTATGATGAGAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATG 154
QY 251 CAGGAGTCCCTTCAGAGATGCGATGCGAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAG 310
DB 155 ATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 214
QY 311 CCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 370
DB 215 AAGATGAAGATGAAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATG 274
QY 371 ACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 425
DB 275 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 329

RESULT 12

US-09-864-761-28059/c
Sequence 28059, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 28059
LENGTH: 327
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005668.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-28059

Query Match 2.4%; Score 43.6; DB 10; Length 327;
Best Local Similarity 47.2%; Pred. No. 0.023;
Matches 133; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 321 TGGTACCCAGAGATGACAGACGCTTATGATGATGATGATGATGATGATGATGATG 380
DB 306 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 247
QY 381 TAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 440
DB 246 CAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 187
QY 441 TACGCTCTACGGAGTAATGATCAAGATCCTTACCTTACCTGAAAGATACAGAACATA 500
DB 186 CAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127
QY 501 TGAACGTGAAGATTCTTGATTAAGCCAGTGATTAATCTTATAGTTTGGCCGAGCTGA 560
DB 126 TGACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 67
QY 561 ACAGACCAAGTGAATTAGAGTGCATGTTTATATCAAGA 602
DB 66 TGATGATGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 25

RESULT 13

US-09-864-761-20733
Sequence 20733, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1

```

CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 20733
LENGTH: 575
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005822.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 13
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 11
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 11
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 12
OTHER INFORMATION: NT HIT: AF185255.1, EVALUE 4.80e+00
OTHER INFORMATION: EST_HUMAN HIT: BE071749.1, EVALUE 5.80e+00
US-09-864-761-20733

```

[illegible]

D_b 401 AGAGAAAGGAAAAGGAGAAGGAGAGAGAGAAGGACCAAGAAAGGAAGAGA 460
Q_y 333 GGATGACAGAGCGCTTGATATGATGACTGGCTGACTACGACTTAGATATAATTGTGA 392
D_b 461 GGAAGAAAGAGAAAGAAAGAGAGAAGGAAGAAAGAGAAGAAAGGAAGAGAGA 520
Q_y 393 GGAAGTGACCAGATCTGAGACTC 418
D_b 521 GGAGCTGCAGCTTGCACCTGAGCAC 546

```

RESULT 14
US-09-864-761-3972
: Sequence 3972, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
:   APPLICANT: Penn, Sharron G.
:   APPLICANT: Rank, David R.
:   APPLICANT: Hanzel, David K.
:   APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aecmlca-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
:   CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
:   PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
:   PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
:   PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
:   PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
:   PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
:   PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
:   PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
:   PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
:   PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
: SEQ ID NO 3972
: LENGTH: 1969
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC005822.1
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 13
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11

```


OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 11
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 14
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.8
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 12
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.4
 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 11
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 12
 US-09-864-761-3972

Query Match 2.4%; Score 43.6; DB 10; Length 1969;
 Best Local Similarity 47.7%; Pred. No. 0.07; Mismatches 139; Indels 0; Gaps 0;
 Matches 127; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 153 AGACAGTAGAGCTGATGAAGAGAGTAAACGCTTATGCTGAGCCAAAGAGAA 212
 DB 623 AG 682
 QY 213 ATTGCAAGAGAGAGTGTGCTGCTGATGAGAGAGAGAGAGAGAGAGAG 272
 DB 683 GAAG 742
 QY 273 CATGACAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 332
 DB 743 AG 802
 QY 333 GGATGACAGGACGCTTGATGATGATGATGATGATGATGATGATGATGATG 392
 DB 803 CGAG 862
 QY 393 GGAAGGTGACCCAGATGCTGAGACTC 418
 DB 863 GGAAGCTGACTTGACACTGAGACAC 888

RESULT 15
 US-10-101-487-74
 Sequence 74, Application US/10101487
 Patent No. US20020169125A1
 GENERAL INFORMATION:
 APPLICANT: LEUNG, DAVID W.
 APPLICANT: BERGMAN, PHILIP A.
 APPLICANT: LOFOUIST, ALAN
 APPLICANT: PIETZ, GREGORY E.
 APPLICANT: TOMPKINS, CHRISTOPHER K.
 APPLICANT: MAGGONER JR., DAVID W.
 TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
 TITLE OF INVENTION: THEREOF
 FILE REFERENCE: 077319/0329
 CURRENT APPLICATION NUMBER: US/10/101,487
 CURRENT FILING DATE: 2002-03-20
 PRIOR APPLICATION NUMBER: 60/277,705
 PRIOR FILING DATE: 2001-03-21
 NUMBER OF SEQ ID NOS: 116
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 74
 LENGTH: 720
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 OTHER INFORMATION: oligonucleotide
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (2)..(718)
 US-10-101-487-74

Query Match 2.3%; Score 43.4; DB 9; Length 720;
 Best Local Similarity 47.9%; Pred. No. 0.042;
 Matches 125; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 154 GACAGGTAGAGCTGATGAAGAGAGTAAACGCTTATGCTGAGCCAAAGAGAA 213
 DB 158 GAAG 217

QY 214 TTGCAAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 273
 DB 218 GAGGAG 277
 QY 274 ATGCAAGATGACGACCGACCGACCGACCGACCGACCGACCGACCGAC 333
 DB 278 GAAG 337
 QY 334 GATGACAGAGCGCTTGATGATGATGATGATGATGATGATGATGATGATG 393
 DB 338 GAGGAG 397
 QY 394 GAAGTGACCCAGATGCTGAG 414
 DB 398 GAAAGAGAGAGAGAGAGAGAG 418

Search completed: March 19, 2003, 21:24:48
 Job time : 474 secs

us-09-315-355a-47.rge

SUMMARIES

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Query Score	Match Length	DB ID	Description
------------	-------------	--------------	-------	-------------

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Description

Detaching C and H

JOURNAL Gene 151 (1-2), 291-296 (1994)
MEDLINE 95129878
PUBMED 7828893
FEATURES

Source Location/Qualifiers

1.1853
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="MRC-5 V2"
/cell_type="fibroblast"
/cfeature_1db="lambda ZAP11"
88.1593

/note="nuclear phosphoprotein (similarity to Saccharomyces cerevisiae Pwp1 protein)"

/product="IEF SSP 9502"
/protein_id="AA65201.1"
/db_xref="GI:17765"

/translation="MNRROYTCVAVRCGVAKETPKVELSEKVEKRLIAEAKELQ
EEGGSDSEETGSPSEDGMSARQARPRELEDGEDDEDTLDDDELAAYDLKDE
EGDPAETLESLGLTVYNSNDPVTTLKDEYEREDFLKPSNDLYVGRABED
OCNLEHYVNOEDSEYVHHDILSLAYPLSEWLNPDSPDSTGNYIAVGNMTPVE
WMDLIDVDSLEPVTFLSKLSKKKKKSSSSEGHDAVLDSMVKLIRNYLAS
ADNTVILMDSLGPAASLAVTDKVTQLQFHPPEAQTLLSGYDKSVALYDRSPDE
SHRMRFSGQIERVTMNFSPCHFLASTDDGFPYVNDABSDKXIFLTMANDETSGID
LSSQIKGLVTASADKYVKIMDILGDRPSLVHSRDMGCVLFSSCCPDPLFIYAFGG
QKGLRWMDISTVSVNEARGRRLVLSARNSISGPPGSSSDTPMES"
1817.1822

polya_signal

polya_site 1853 368 c 461 g 496 t

BASE COUNT 528 a 368 c 461 g 496 t

Query Match 100.0%; Score 1853; DB 9; Length 1853;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1853; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCCCTGACGCTGTGACACAGTGCCTGTGCTCCCTCTATGACGCTGGTTCTA 60
Db 1 GATCCCTGACGCTGTGACACAGTGCCTGTGCTCCCTCTATGACGCTGGTTCTA 60
QY 61 GCCTGACAGCCCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 120
Db 61 GCCTGACAGCCCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 120
QY 121 TGGGTCGCTGCGCGCTGCGCAAGAGACACAGCAAGGTAGAGTGAAGTGAAGTGA 180
Db 121 TGGGTCGCTGCGCGCTGCGCAAGAGACACAGCAAGGTAGAGTGAAGTGAAGTGA 180
QY 181 GTAAAACGCTCTATTGCTGAGGCAAGAAATTGCAAGAAAGTGTGTGAGTGTAT 240
Db 181 GTAAAACGCTCTATTGCTGAGGCAAGAAATTGCAAGAAAGTGTGTGAGTGTAT 240
QY 241 GAAGAGAGACAGGAGTGTCTCAGAAAGTGTGACAGAGTGCAGCCAGCAGCAGC 300
Db 241 GAAGAGAGACAGGAGTGTCTCAGAAAGTGTGACAGAGTGCAGCCAGCAGCAGC 300
QY 301 CCAAGAGAGCCCTGAGAGTGTGACCCAGAGATGACAGAGCTTGAATGATGAG 360
Db 301 CCAAGAGAGCCCTGAGAGTGTGACCCAGAGATGACAGAGCTTGAATGATGAG 360
QY 361 CTGGCTGAGTACGATTGATTAATATGATGAGAAAGTGAAGTGAAGTGAAGTGA 420
Db 361 CTGGCTGAGTACGATTGATTAATATGATGAGAAAGTGAAGTGAAGTGAAGTGA 420
QY 421 GGTGAATCTCTTGGGCTTACGCTTACGAGAGTAAATGATCAAGATCTTACGTA 480
Db 421 GGTGAATCTCTTGGGCTTACGCTTACGAGAGTAAATGATCAAGATCTTACGTA 480
QY 481 CTGAAGATTAAGAAATATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 540
Db 481 CTGAAGATTAAGAAATATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 540
QY 541 ATAGTTGTGCGCAGAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 600
Db 541 ATAGTTGTGCGCAGAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 600

Db 541 ATAGTTGTGCGCAGAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 600
QY 601 GAAGAGAGTCTTTTATATGATACACATGATATATCTTTGCTGATATCTCTGAGTGT 660
Db 601 GAAGAGAGTCTTTTATATGATACACATGATATATCTTTGCTGATATCTCTGAGTGT 660
QY 661 GAATGGCTGAATTTGATATCTTACGAGCCAGATGATCTTCTGAAATATCATTTGCTGATGA 720
Db 661 GAATGGCTGAATTTGATATCTTACGAGCCAGATGATCTTCTGAAATATCATTTGCTGATGA 720
QY 721 AACATGACCCCTGTTATGATGAGTGGAGCTGATATATGATGATCTTTAGGCCAGTCT 780
Db 721 AACATGACCCCTGTTATGATGAGTGGAGCTGATATATGATGATCTTTAGGCCAGTCT 780
QY 781 TTCACTCTGGAAGTAACTTTTCAAAAAAGAGAGAGAGAGAGAGAGAGTCTCTCA 840
Db 781 TTCACTCTGGAAGTAACTTTTCAAAAAAGAGAGAGAGAGAGAGAGTCTCTCA 840
QY 841 GCAGAGAGGATACCGATGCTGCTTGAACCTTTCATGCAATGATATGCAAAATGTT 900
Db 841 GCAGAGAGGATACCGATGCTGCTTGAACCTTTCATGCAATGATATGCAAAATGTT 900
QY 901 TTAGCAAGTGCATCAGCTGACAACTGTAATCTGTGGATATGCTTGGGGAACCA 960
Db 901 TTAGCAAGTGCATCAGCTGACAACTGTAATCTGTGGATATGCTTGGGGAACCA 960
QY 961 GCAGCTAGCCCTGCTGTATACACAGAGCAAGTTCACCACTGCACTTTCATCTTGA 1020
Db 961 GCAGCTAGCCCTGCTGTATACACAGAGCAAGTTCACCACTGCACTTTCATCTTGA 1020
QY 1021 GCACAGACTCTGATTTTGTGCTCATATGATATGATGATGATGATGATGATGATGAT 1080
Db 1021 GCACAGACTCTGATTTTGTGCTCATATGATATGATGATGATGATGATGATGATGAT 1080
QY 1081 CCAAGATGAAGCCATGCAATGTGGGATTCAGTGGGAGATGAGAGTCACTTGAT 1140
Db 1081 CCAAGATGAAGCCATGCAATGTGGGATTCAGTGGGAGATGAGAGTCACTTGAT 1140
QY 1141 CACTTTTCACTTTCATTTTGTGGCCAGTACATGATGATGATGATGATGATGATGAT 1200
Db 1141 CACTTTTCACTTTCATTTTGTGGCCAGTACATGATGATGATGATGATGATGATGAT 1200
QY 1201 GCACGTTAGATTAAGCCATTTTACCTTATGACACATGATGATGATGATGATGAT 1260
Db 1201 GCACGTTAGATTAAGCCATTTTACCTTATGACACATGATGATGATGATGATGAT 1260
QY 1261 GATCTTACAGTCAATCAATCAAGGCTGTCTGTGACTCTTCACTGACCAATATGATGAG 1320
Db 1261 GATCTTACAGTCAATCAATCAAGGCTGTCTGTGACTCTTCACTGACCAATATGATGAG 1320
QY 1321 ATCTGGACATCTTAGAGATAGGCAAGTCTGTGATGATGATGATGATGATGATGATGAT 1380
Db 1321 ATCTGGACATCTTAGAGATAGGCAAGTCTGTGATGATGATGATGATGATGATGATGAT 1380
QY 1381 GTTCTCTTCTGTTCTTCAATGCTGATGATGATGATGATGATGATGATGATGATGAT 1440
Db 1381 GTTCTCTTCTGTTCTTCAATGCTGATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 AAAGAGAGGCTTGGGCTGAGGATTAAGACAGTCTTCAATTAATGAAGATTTGGA 1500
Db 1441 AAAGAGAGGCTTGGGCTGAGGATTAAGACAGTCTTCAATTAATGAAGATTTGGA 1500
QY 1501 AGAGAGAGAGGCTTGTCTTGGAGTGAAGAAATTCATATGATGAGCCCTTTTGGC 1560
Db 1501 AGAGAGAGAGGCTTGTCTTGGAGTGAAGAAATTCATATGATGAGCCCTTTTGGC 1560
QY 1561 AGAGAGAGTCAATACACCATGAGTCTTATGAGATCAATTAATTTCTGCTTACC 1620
Db 1561 AGAGAGAGTCAATACACCATGAGTCTTATGAGATCAATTAATTTCTGCTTACC 1620
QY 1621 TTAACTGGGAATTTTAAAGTTGGCTAAAGTTCATGATGCTGGCAGCAACATGCA 1680
Db 1621 TTAACTGGGAATTTTAAAGTTGGCTAAAGTTCATGATGCTGGCAGCAACATGCA 1680

Qy	219	AGAGAAAGTGTGTGCACTGATGTAAGAAGAGACAGGACAGTCCCTTACGAAGATGGCATTGCA	278
Db	194	AGAAAGAA--GGTGGAGGTGAAGAAAGAGAGGACAGCAACCTTACGAAGATGGCATGCA	250
Qy	279	GAGTGCACGCAACCGACGACGCCCAAGAGAGCCCTCGAGAGATGGTGAACCCAGAGATGA	338
Db	251	GAGTGGCCCGAACCCAGGACACCGCTTGGAGATGTCCTCGAAGATGGGGATCCACAAGATGA	310
Qy	339	CAGGACGCTTGATGATGATGATGAGCTGGCTGATGACGATTTAGATAAATATGATGAGGAAG	398
Db	311	CCGCACTGAGACGATGACGATGAGCTGGCGGGATACGACTTATATTTATGACGAAGAGGA	370
Qy	399	TGACCCAGATGCTGAGACTTGGTGAAATCTCTTTGGGTCTTACGGTCTACGGGAGTAA	458
Db	371	TAAACCCAGATGCTGAAGACCAATGGCAAGTCTGTGGGGTCTTAAGGCTATGGCAGTAA	430
Qy	459	TGATCAAGATCCTTAAGTCTCTGAAAGATACAGAAACAATATGAAAGTGAAGATTTCTT	518
Db	431	TGACCAAGATTCATATGTCACTTGAAAGACACGGAACAATATGACATGAAGATTTCTT	490
Qy	519	GATTAAACCCAGATGATATCTTATTAAGTTTGTGGCCAGCTGAACAGACAGTGCATTTT	578
Db	491	GATTAAACCGACATGACAACTCATCGTTGTGGCCGTGCTGAACAAGACAGTGCATCT	550
Qy	579	AGAGTGCATGTTTATATATCAAGAAAGAACTCTTTTATATGACACATGATATCTTT	638
Db	551	AGAGGTCAAGTTTACATATCAAGAAAGAGAGTCTTTCTAGTCCACATGACATATCTCT	610
Qy	639	GTCGCAATCCCTGAGTGTGGAATGGCTGAATTTGATCCTTACCCAGATGATTCAC	698
Db	611	GTCGCAATCCCTGTGAGTGTGAATGGCTGAATTTGATCCTTACCCAGATGATTCAC	670
Qy	699	TGCAAAATTAATTCCTGTGTAGAGAAACATGACCCCTGTATATGAAGTGTGGACCTTGATAT	758
Db	671	TGGGAATTAATTCCTGTGTGGGAACATGATCCTGTTCATCGAAGTATGGGACTGGATAT	730
Qy	759	AGTGAATCTTTTAGAGGCACTCTTCACTCGAGATGAATCTTTCAAAAAAGAGAAAA	818
Db	731	AGTGAATCTCTAGAAACCACTCTTCAACCTGGAGACAAGCTTTCCAAAAAGAGAAAA	790
Qy	819	GAAAGGAAAGAAAGTTCCTCAGCAGAAAGGCAATCCGATGCTGTCCTTGACCTTCATG	878
Db	791	GAAAGGAAAGAAAGTTCCTCAGCAGAAAGCAATCTGATGCTGTCTTGAATATCTCG	850
Qy	879	GAAATAGCTATCGAAATGTTTTTAGCAAGTGCATCAGCTGACCAACACTGTAAATCTGTG	938
Db	851	GAAATAGCACTCGAAATGTCTGTCAAGTGCATGCGACAGACACTGTATGTTCTCTG	910
Qy	939	GGATATGTCCTTGGGGGAAACCAAGACAGTACGCTCGCTGACACAGACAAGGTCCAAC	998
Db	911	GGATCTGTCTGTGGGAAAGTCAGTAGCAGGCTTACTGCACATACAGACAAGGTTCAAC	970
Qy	999	ACTGCAGTTTCACTCAATTTAAGACAGACTGTGATTTCTGGCTCATATGATATGTCAGT	1058
Db	971	TCTCAGATTTCACTCAATTTAAGACAGACTGTGATTTCTGGGCTCTATGACAATACTCGT	1030
Qy	1059	GGCTTTGATGACCGCCGAAGTCCAGATGAAAGCAATGCAATATGTGGCCATTCAGTGGGCA	1118
Db	1031	GGCTCTGTATGATCGCAGAAATCCAGCCGAACAACCCGCAAGTGGCGGTTAGTGGGCA	1090
Qy	1119	GATAGAGAGATGACTTGGAGATCACTTTTCACTTGTATCTTTCTTGGCCAGTACAGATGA	1178
Db	1091	GATGAGAGATGACTGTGAACCACTTCTCAACCTGTATCTTGGCCAGTACAGATGA	1150
Qy	1179	CGGCTTTGTATATTAATTTGATGACGTTCAATTAAGCCAATTTTTCACCTTAATGACA	1238
Db	1151	TGGCTTTGTATATTAACCTGTGATGACGTTTCAACAAGCCAATCTTTACCTTAATGACACA	1210
Qy	1239	CAATGATGAATCTCTGGCTGTATCTTACACATCAAAATCAAGGGCTGTCTGTATACGC	1298
Db	1211	CAACATGAATATCTCTGGCTGTATCTTACACATCAAAAGGCTGTCTGTATACGC	1270

Oy	1239	TTGGCTGACAAATAACGGAAGATCTGGGACATCTTAGAGATGAGCCAAAGTCA	1358
Db	1271	TTCCGCTCAAAATTTTGAAGATCTGGGACATCTTGAAGATTAGCCAAAGTCTCATTC	1330
Oy	1359	TTCTAGGACATGAAAAATGGGAGTTCCTTCTGTCTTCATGTTGCCCTGATTTGCCATT	1418
Db	1331	CTCTAGACACATGAAAAATGGGAGTTCCTTTTGTTCCTTCCTGTCTCTGACTTGCACATT	1390
Oy	1419	TATTTATGCTTTGGAGGTCAAAAAGANGGCTTGGGCTCTGGGATATTAAGCACAGTCTC	1478
Db	1391	TGTTATATCCTTTGGAGGACAGAAAAGAGGCTTCGGGTTTGGGACATTAACACATGTGTTC	1450
Oy	1479	TTCACTAAATGAGACATTTGGAAAGCGAGAGGCTGTCTTGGGAGTGCAGAAATTC	1538
Db	1451	TTCACTAAATGAGACATTTGGAAAGCGGAGGACACTCCTTATAGCGGACAGAAAGTTT	1510
Oy	1539	ATCTATTAGTGCCCTTTTGGCAGCAGAGCTCA--GATACACCCATGAGTCTTAATG	1595
Db	1511	ATCTGTTAGCGGCCCTCTGGGAGCAGGAGTCCACAGCAGACCAATGGAGTCTTGACA	1570
Oy	1586	AAGATCA-TCCTAATTTCTGCTTACCTTAACCTGGGAATTTTAAAAAGTTGGCTTAAAT	1654
Db	1571	AAGATCAGTTTCCTGTGTGCTTACCTTAACCTCG--TTTCAAAAAGACTGGGCTTAAAAAT	1628
Oy	1655	GTTTC	1658
Db	1629	GTTCC	1632

LOCUS	BC003199	1828 bp	mRNA	linear	ROD 07-AUG-2002
DEFINITION	Mus musculus, RIKEN cDNA 2310058A11 gene, clone MGC:6406				
IMAGE	3586496, mRNA, complete cds.				
ACCESSION	BC003199				
VERSION	BC003199.1	GI:13096811			
KEYWORDS	MGC.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 1828)
Strausberg, R.
Direct Submission
Submitted (20-FEB-2001) National Institutes of Health, Mammalian

REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLB at: <http://image.llnl.gov>
Series: IRAK Plate: 10 Row: 6 Column: 6
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer Frequency ORF analysis, Genomescan gene prediction, similarly but not identically to protein.

FEATURES

Location/Qualifiers

```
source      1. .1828
/organism="Mus musculus"
```

```

/db_xref="taxon:10090"
/map="CZECH 11"
/clone="MGC:6406 IMAGE:3586496"
/issue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clone) outgrowth infected with the virus MMTV."
/clone_1db="NCI CGAP_Lu29"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
82..1587
/codon_start=1
/product="RIKEN CDNA 2310058A11 gene"
/protein_id="AAH03199.1"
/db_xref="GI:13096812"
/translation="MNRSRQVTCVAVWRGVAKETPDKVELSKVEVRLIABAKKXLO
EEGSEBEEAGNPSDEDMQGPAPRESLEDDPDQDRLDDDELADGLDNLNDE
DNDAETIGSLGLTVYGSNDODPYTLDTQTEYHEHDLIPIPTDLIVCGABEOB
CNEFHYHNOBESFYHHDILSAYPLSEYINAFDPSPASTGNTIAYGNMTPIUEV
WDLDVLSLPVFTLSKLSKKKKKKSSABGMDAVALDLSMNTVNLVLSASA
DSTVVLMDLSVGRSVARLTATDKVQTLQHPPEAQLTIGSDVKSVALYDCRDPSON
HRQMFSGQIERVTMNFSPCHFLASTDDFVYNLDARSDKPIFTLAAHNDLSGLDL
SSQIKGLTVASADKLVKIDILGDRPSLIHSRDMKKVGFCSGCCDLPFFVYAFGGQ
KEGRVMDISTVSVNEAFGRRLVIGTKLSVSGSRSPQOTPMES"
BASE COUNT      504 a      414 c      476 g      434 t
ORIGIN
Query Match      62.2%; Score 1152.4; DB 10; Length 1826;
Best Local Similarity 82.7%; Pred. No. 2e-102;
Matches 1367; Conservative 0; Mismatches 276; Indels 9; Gaps 4;
QY 11 CGTGTGCAGAGTCGGTCTGTCTCCCTCCCTATGACAGCTGTGTTTACGTCGACAG 70
Db 5 CGCCTGGCGCGGAGCTCAAGAGTAGTGTCTTCCAGCTCCAGGAGCTTCAACCTGCGCG 64
QY 71 CCCTTGACTGAGGACCATGACCGCAGCGCCAGGTGACGTCGCGCTGGGCTCGCT 130
Db 65 GGTTCMAACCAAGATCTGAACCGCACGCCCAAGTACCTGGCTGGCTCGGCTCGCC 124
QY 131 GCGGCGTGGCCAAAGAGACACACAGACAGTAGAGCTGAGTAAAGAAAGTAAACGCC 190
Db 125 GCGGCGTGGCCAAAGAGACCTCAAGCAAGTAGAGCTAAGTAAAGAGAACTGAACCGTC 184
QY 191 TCATTGCTGAGGCAAGAGAAATTGCAAGAAAGGTGTGCGAGCTGATGAAGAGAGA 250
Db 185 TCATCGCTGAAGCAAGAAAGTTCAGAAAGAA---GATGGAGTGAAGAAAGAGAGG 241
QY 251 CAGGCAAGCTTCAAGAGATGAGCATGAGATGACCGCACCCAGGCAAGGCCAAGAGAGC 310
Db 242 CAGGCAAGCTTCAAGAGATGAGCATGAGATGAGGCGCATCAAGGCAAGCGCTTAGGAGT 301
QY 311 CCCTGAGAGATGTGACCCAGAGAGTGAACAGACGCTTGAATGATGATGAGCTGAGT 370
Db 302 CCCTGAGAGATGAGGAGTCAACAAGTGAACGCAACATGAGAGATGAGAGAGAGGAGT 361
QY 371 ACGACTTGAATTAATGATGAGAGAGAGTGAACCAAGATGCTGAGACTTTGTGATATTC 430
Db 362 ACGACTTGAATTAATGATGAGAGAGATGAACCAAGATGCTGAGAGCAATTCGCAAGTCTC 421
QY 431 TCTTGGGCTTGAAGGTCTAAGGAGATATGATCAAGATCCTTAGCTTACTGTAAGATA 490
Db 422 TGTGGGCTTGAAGGTCTTATGAGAGATATGACAGAGATCCATATGTCACCTTGAAGACA 481
QY 491 CAGAACATATGAACGTGAAGATTTCTGATTAAGCCCAAGTATATCTTAAGTTGTG 550
Db 482 CAGAACATATGAACGTGAAGATTTCTGATTAAGCCCAAGTATATCTTAAGTTGTG 541
QY 551 GCCGAGCTGAACAGAGACAGTGAATTTAGAGGTGATGTTTATATATGAAGAAAGACT 610
Db 542 GCCGTGCTGAACAGAGACAGTGAATTTAGAGGTGATGTTTATATCAACAAAGAGAGT 601
QY 611 CTTTATATGACACATGATATACCTTGTCTGCAATATCCTCTGAGTGTGAATGGCTGA 670

```

```

Db 602 CTTTCTAGCTCAACATGACATATCTCTGTCTGATACCTCTGAGTGTGATGGCTGA 661
QY 671 ATTTGATCTTAAGCCAGATGATGATTTACTGGAATTAATCATTTGCTGTAGAAACATGACC 730
Db 662 ATTTGATCTTAAGCCAGATGATGATTTACTGGAATTAATCATTTGCTGTGAGAAACATGACC 721
QY 731 CTGTATTTGAAGTGTGGACCTTGAATATAGTGAATCTTTAGAGCCAGTCTTCACTG 790
Db 722 CTGTATTTGAAGTGTGGACCTTGAATATAGTGAATCTTTAGAGCCAGTCTTCACTG 781
QY 791 GAAATTAATCTTCAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 850
Db 782 GAAGCAAGCTTCCAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 841
QY 851 ATACCAATGCTGTCTTGAACCTTATGATGAATGATGATGATGATGATGATGATGATG 910
Db 842 ATACCAATGCTGTCTTGAACCTTATGATGAATGATGATGATGATGATGATGATGATG 901
QY 911 ATACCAATGCTGTCTTGAACCTTATGATGAATGATGATGATGATGATGATGATGATG 970
Db 902 ATACCAATGCTGTCTTGAACCTTATGATGAATGATGATGATGATGATGATGATGATG 961
QY 971 TCGCTGACACAGACAGCAAGTCCAAACACTGCAAGTTTCATCCATTTGAAGCAGAGACTC 1030
Db 962 TTAATGACATATGACAGCAAGTCCAAACACTGCAAGTTTCATCCATTTGAAGCAGAGACTC 1021
QY 1031 TGATTTCTGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1090
Db 1022 TGATTTCTGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1081
QY 1091 GCAATGCAATGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1150
Db 1082 ACCACCGCAGAGCGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1141
QY 1151 CTGTGATTTCTTGGCCAGTACAGATGACCGCTTGTATATTAATTTGATGATGATGATGATG 1210
Db 1142 CGTGTATTTCTTGGCCAGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1201
QY 1211 ATAAAGCAATTTTACACTTAATGACACAAATGATGATGATGATGATGATGATGATGATG 1270
Db 1202 ACAAGCAATTTTACACTTAATGACACAAATGATGATGATGATGATGATGATGATGATG 1261
QY 1271 GTCAATCAAGGGGCTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1330
Db 1262 GTCAATCAAGGGGCTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1321
QY 1331 TCTTGAAGATAGGCGCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1390
Db 1322 TCTTGAAGATAGGCGCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1381
QY 1391 GTTCTTCAATGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1450
Db 1382 GTTCTTCAATGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1441
QY 1451 TTGGGCTGTGGGATATTAAGACAGTCTTCAATATGAAGATTTGGAAGACGAGAGA 1510
Db 1442 TTGGGCTGTGGGATATTAAGACAGTCTTCAATATGAAGATTTGGAAGACGAGAGA 1501
QY 1511 GGCCTGTTCTTGGGAGTGAAGAAATTCATCTATTAAGTGGCCCTTTTGGCAGCAGAGACT 1570
Db 1502 GACTCGTTATAGCGGAGAGAAAGTTTATCTGTATGCGGCCCTGTGGGAGCAGAGAGT 1561
QY 1571 CA--GATACACCATGAGACTTTAATGAAGATCA--TCTAATTTCTGCTTACTTACT 1626
Db 1562 CACAGCAGACCAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1621
QY 1627 GGGATTTTAAAGTGTGGCTTAAATGTTTC 1658
Db 1622 CCG--TTTCAAAAGACTGGCTTAAATGTTTC 1651

```

RESULT 5
AC098211/c

LOCUS AC098211 164664 bp DNA linear HTG 12-JUL-2002
DEFINITION Rattus norvegicus clone CH230-208L16, *** SEQUENCING IN PROGRESS
ACCESSION AC098211
VERSION AC098211.3 GI:21729861
KEYWORDS HTG; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 164664)
Munzy,D.M., Adams,C., Adio-Ochola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Birmage,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durdin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,T., Foster,P., Frantz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Hornel,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kritovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,X., Liu,R., Louisged,H.,
Lozdo,R.V., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,N., Morris,S.,
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunnu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojokokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanil,K., Vaequez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 164664)
REFERENCE Worley,K.C.
JOURNAL Direct Submission
TITLE Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 164664)
REFERENCE Worley,K.C.
JOURNAL Direct Submission
TITLE Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17973812.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: GIMR
Center clone name: CH230-208L16
----- Summary Statistics
Sequencing Vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 106583 bases at least Q40
Consensus quality: 113107 bases at least Q30
Consensus quality: 118472 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 113: contig of 1313 bp in length
* 1314 1413: gap of unknown length
* 1414 2511: contig of 1098 bp in length
* 2512 2611: gap of unknown length
* 2612 3846: contig of 1235 bp in length
* 3847 3946: gap of unknown length
* 3947 5519: contig of 1573 bp in length
* 5520 5619: gap of unknown length
* 5620 7133: contig of 1514 bp in length
* 7134 7233: gap of unknown length
* 7234 8594: contig of 1361 bp in length
* 8595 8694: gap of unknown length
* 8695 9703: contig of 1009 bp in length
* 9704 9803: gap of unknown length
* 9804 10945: contig of 1142 bp in length
* 10946 11045: gap of unknown length
* 11046 12242: contig of 1197 bp in length
* 12243 12342: gap of unknown length
* 12343 13881: contig of 1539 bp in length
* 13882 13981: gap of unknown length
* 13982 15230: contig of 1249 bp in length
* 15231 15330: gap of unknown length
* 15331 16364: contig of 1034 bp in length
* 16365 16464: gap of unknown length
* 16465 18075: contig of 1611 bp in length
* 18076 18175: gap of unknown length
* 18176 19710: contig of 1535 bp in length
* 19711 19810: gap of unknown length
* 19811 21291: contig of 1481 bp in length
* 21292 21391: gap of unknown length
* 21392 22816: contig of 1425 bp in length
* 22817 22916: gap of unknown length
* 22917 24259: contig of 1343 bp in length
* 24260 24359: gap of unknown length
* 24360 26369: contig of 2010 bp in length
* 26370 26469: gap of unknown length
* 26470 28190: contig of 1721 bp in length
* 28191 28290: gap of unknown length
* 28291 29566: contig of 1276 bp in length
* 29567 29666: gap of unknown length
* 29667 31464: contig of 1798 bp in length
* 31465 31564: gap of unknown length
* 31565 32733: contig of 1169 bp in length
* 32734 32833: gap of unknown length
* 32834 34340: contig of 1507 bp in length
* 34341 34440: gap of unknown length
* 34441 35861: contig of 1421 bp in length
* 35862 35961: gap of unknown length
* 35962 37213: contig of 1252 bp in length
* 37214 37313: gap of unknown length
* 37314 38899: contig of 1586 bp in length
* 38900 39000 40786: contig of 1787 bp in length

```

* 40787 40886: gap of unknown length
* 40887 42840: contig of 1954 bp in length
* 42841 42940: gap of unknown length
* 42941 44820: contig of 1880 bp in length
* 44821 44920: gap of unknown length
* 44921 46286: contig of 1376 bp in length
* 46287 46397: gap of unknown length
* 46397 49196: contig of 2800 bp in length
* 49197 49296: gap of unknown length
* 49297 51055: contig of 1759 bp in length
* 51056 51155: gap of unknown length
* 51156 52468: contig of 1313 bp in length
* 52469 52568: gap of unknown length
* 52569 54246: contig of 1678 bp in length
* 54247 54346: gap of unknown length
* 54347 56244: contig of 1898 bp in length
* 56245 56344: gap of unknown length
* 56345 57759: contig of 1415 bp in length
* 57760 57859: gap of unknown length
* 57860 59092: contig of 1233 bp in length
* 59093 59192: gap of unknown length
* 59193 62231: contig of 3039 bp in length
* 62232 62331: gap of unknown length
* 62332 65033: contig of 2702 bp in length
* 65034 65133: gap of unknown length
* 65134 67535: contig of 2402 bp in length
* 67536 67635: gap of unknown length
* 67636 70030: contig of 2395 bp in length
* 70031 70130: gap of unknown length
* 70131 72994: contig of 2864 bp in length
* 72995 73094: gap of unknown length
* 73095 75682: contig of 2588 bp in length
* 75683 75782: gap of unknown length
* 75783 78533: contig of 2750 bp in length
* 78534 78633: gap of unknown length
* 78634 81157: contig of 2525 bp in length
* 81158 81257: gap of unknown length
* 81258 83896: contig of 2639 bp in length
* 83897 83996: gap of unknown length
* 83997 86151: contig of 2155 bp in length
* 86152 86251: gap of unknown length
* 86252 88033: contig of 2782 bp in length
* 88034 89133: gap of unknown length
* 89134 92970: contig of 3837 bp in length
* 92971 93070: gap of unknown length
* 93071 95027: contig of 1957 bp in length
* 95028 95127: gap of unknown length
* 95128 98206: contig of 3079 bp in length
* 98207 98306: gap of unknown length
* 98307 100157: contig of 1851 bp in length
* 100158 100257: gap of unknown length
* 100258 103213: contig of 2956 bp in length
* 103214 103313: gap of unknown length
* 103314 107610: contig of 4297 bp in length

Query Match      53.2%; Score 986.6; DB 2; Length 164664;
Best Local Similarity 79.8%; Pred. No. 5.3e-257;
Matches 1257; Conservative 0; Mismatches 269; Indels 49; Gaps 6;

```

```

Db 157160 TTCAGAAAGATG----GCAGAGACCCGGAGCCAGGTCCAGTCTAGGAGAGCACTGGAGGA 157105
Qy 321 TGTGATCCAGAGATGACGAGACGCTTGATGATGATGAGCTGGCTGAGTACGACTTTAGA 380
Db 157104 TGGGGATCCCAAGACAGACACACACTGGACGACGATGAATCGGCCAAGTACAC----- 157050
Qy 381 TAAATATGATGAGAAAGGTGACCCAGTGTGAGACTCTGGTGAATCTCTCTGGGCT 440
Db 157049 -----GAAGAGACAAACCAAGTGTGAGACCATTTGGTGAATCTCTGTGGGTCT 157000
Qy 441 TACGTTACGGGAGTATGATGATCAAGTCTTCTTACGTTACTCTGAAAGATACAGAAACATA 500
Db 156999 TACTGTAGGGAGTATGATGAGGTGACGCTGACATCAATCAAGAAAGAGAGTCTTTCTATGT 156940
Qy 501 TGAACGTGAAGATTTCTTGATTTAAGCCAGTGAATCTTATAGTTTGTGCGGAGCTGA 560
Db 156939 TGAGCGTAGAAGACTCCTTGATTTAAACGACCTGGCAACCTCATGCTGTGGCGGTGTGA 156880
Qy 561 ACAGGACAGTGCATTTAAGGTCATGTTTAAATCAAGAAAGAACTCTTTTATGT 620
Db 156879 ACAAGGCGAGTGTACTTGAAGGTGACGCTGACATCAATCAAGAAAGAGAGTCTTTCTATGT 156820
Qy 621 ACACATGATATACTCTTGTCTGCATATCCTCTGAGTGTGAATGCTGAATTTGTATCC 680
Db 156819 CCACATGACATCTCTCTGTCTGATATC--CTGAGTGTGAATGCTGAATTTGTATCC 156762
Qy 681 TAGCCAGATGATTTACTGAAATTAATTGCTGTGTAAGAAACATGACCCCTGTATTTGA 740
Db 156761 TAGCCAGATGCGGCTACTGAAATTAATTGCTGTGTAAGAAACATGACCCCTGTATTTGA 156702
Qy 741 AGCTGGACCTTGATATGATGACCTTTAAGCCAGTCTTCACTCGGAAGTAACT 800
Db 156701 GGGGTGGACCTGATATCGGACCTCTTAAGACAGTCTTCACTCGGAAGTAACT 156642
Qy 801 TTCAAAAAGAGAAAAAGAAAGAAAGAGTCTCTCAAGAGAGGACATACCATG 860
Db 156641 TTCAAAAAGAGAAAGAAAGAAAGAGTCTCTCAAGAGAGGACATACCATG 156582
Qy 861 TGTCTTGACCTTCATGAAATTAAGTAACTGTAAGAGTCTTGAAGTCAATGCTATGCTGA 920
Db 156581 TGTCTTGACCTTCCTGAGTAAAGACAGTCAAGAAAGTCTGTGTGATGATAGAGA 156522
Qy 921 CAACAGTAAATCTGT-----GGGATATGTCTTGGGAAACCGAGCTAGCTGCTG 976
Db 156521 CAGACGTAAATCTGTGGGGGAGCTGCTGTGGGAAAGTCAAGCCAGGTTTACTG 156462
Qy 977 TACACAGACAAAGTCAAAACACTGCAAGTTTCAATTCATTTGAAGCAGACCTGATTT 1036
Db 156461 CACACAGACGAGAGTTAGCTGTGCAAGTTTATTCATTTGAAGCAGACCTGATTT 156402
Qy 1037 CTGCTCATATGATTAAGTCAAGTGTCTTTGTAATGACCTGCGAAGTCCAGATGAAGCCATC 1096
Db 156401 CTGGGTTCTATGAATAATCCGTGGCTCTGTAAGCAGTCAAGATCCAGCCAGATATACC 156342
Qy 1097 GAATGTGCCATTTAGTGGGAGATAGAGAGAGTGAATCACTTTTCACTTGTCT 1156
Db 156341 GTCTGTGTGTTAGTGGGAGATAGAGAGCTGACCTGGAACCACTTTCTCACGCTCC 156282
Qy 1157 ATTTCTTGCCAGTACAGATGACCGCTTTTGTATTAATTTGGATGACGTTTCAGATTAAC 1216
Db 156281 ATTTCTTAGCCAGTACAGATATGCTTTGTGTATTAATCTGAGTGGCGGCTC----- 156230
Qy 1217 CAATTTTACATTAATGACACAAATGATGAATCTCTGCTTTGATCTTGAAGCTAA 1276
Db 156229 -----TACTGACACAAAGATGAATCTCTGCTTTGATCTGAGCTAGGA 156183
Qy 1277 TCAAGGCTGTCTGTGACCTTCAAGTGAATAATACGTAAGTACGAGACATCTTGA 1336
Db 156182 TCAAGACTGCTCTTGTGACCTTCAAGCAAAATTTGACATTTGAGACATCTTGA 156123
Qy 1337 GAGATAGGCGCAAGTCTAGTTCAATCTAGGACATGAATAAGGAGTCTCTCTGTTCTT 1396
Db 156122 GGGATAGCAAGTCTATTCATTTAGAGATATGAATAAGGAGTCTCTCTTGTGTTCTT 156063

```

QY	1397	CATGTGGCCCGTCATTTGGCATTTATTTATGCTTTGGAGAGTCATAAAAGAAAGGCGCTTCGGG	1456
Db	156062	CCTGCTGTCCTCATTTATTTGCCCTTTGTTTAAGCGCTTTGGAGACAGAAAGAGGCGCTTCGGG	156003
QY	1457	TCTGGGATATTAAGACAGACAGTCTCTTCAGTAAATGAAGCATTTTGAAGACAGAGAGCTTGG	1516
Db	156002	TTTGGGACATAAACAAGTGTGTCTTCAGTAAAGAGACGCTTTGGAAGACGGGAGAGACTGGG	155943
QY	1517	TTCTTGGGAGTGCAGAAATTCATCTTATTAAGTGGCCCTTTGGAGACAGAGAGCTCANATA	1576
Db	155942	TTATTTGGACAGACGAAAGTTTATCTGTTATTTAGTCCCTCTTGAAGACAGAGACCCACAGA	155883
QY	1577	CACCATGAGAGCTTAAATGAAGATCATCTAAATTTCTGCTTACCTTAATCTGGAATTTTAA	1636
Db	155882	CACCAATGAGAGCTTCGACAGAAAGATAGAGTTCCCTGAGTGTGCTTAACTTGAAGATTAA	155823
QY	1637	AAAAGTTGGCCTTAA 1651	
Db	155822	AAAAGTTGGCCTTAA 155808	
RESULT	6		
BC000067		807 bp	mRNA
LOCUS			linear
DEFINITION			PRI 12-JUL-2001
ACCESSION			
VERSION			
KEYWORDS			
ORGANISM			
SOURCE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			
FEATURES			
Source			
BASE COUNT			
ORIGIN			
Query Match	42.4%	Score 784.8	DB 9; Length 807;
Beet Local Similarity	99.1%	Pred. No. 2.3e-202;	

[illegible]

Alsbrooke, S.L., Amaralunga, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Biewe, K., Blankenburg, K., Bonini, D., Bouck, J., Bowles, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy, Carroll, L., Dederich, D.A., Delaney, K.R., Desigado, O., Dunn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Barnhart, C., Edgar, D., Edwards, C.C., Elms, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, L., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Goirell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Joliver, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mel, G., Merscher, S., Metzger, M., Miller, A., Miner, G., Morris, Z., Mitchell, T., Mohabac, K., Montgomery, K.T., Morgan, N., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, M., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nockenkwo, S., Ogunu, M., Okunolu, G., Otagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pyl, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shih, C., Siohastari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Taney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasequez, L., Verra, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Kucherlapati, R. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 121700)
Worley, K.C.
Direct Submission
Submitted (21-MAY-1999) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 121700)
Worley, K.C.
Direct Submission
Submitted (01-NOV-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 1, 2000 this sequence version replaced gi:11038616.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for human and mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3369-3402) demonstrably (expect < 1e-24) to the EST and cDNA sequences. Genes demonstrate at least two exons

flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 207416
Phrap values in estimate: 206947
Average error rate (BCM-Phrap estimate): 9.47342e-05
Fraction of Phrap values less than 40 : 0.0155888
Number of consensus changing edits: 31
Number of N's in consensus : 0

Position	Consensus changing edits	Edited Context
22181	atggtctcagc(n)ctgtaatccc	atggtctcagc(c)ctgtaatccc
22487	ttttccatgct(n)ngctctgac	ttttccatgct(a)ngctctgac
22488	ttttccatgct(n)ngctctgac	ttttccatgct(a)ngctctgac
22503	tgaaactatc(n)tatatgctt	tgaaactatc(c)tatatgctt
22523	tattccatga(n)gtttatagg	tattccatga(t)gtttatagg
22563	taagatctac(c)ctaatatnaat	taagatctac(a)ctaatatnaat
22666	gattctacc(n)atnaatgac	gattctacc(g)atnaatgac
22678	ctaccatnaat(n)aatgacacm	ctaccatnaat(a)aatgacacm
22679	naaatgacac(n)naatgacacm	naaatgacac(g)naatgacacm
22683	tgcaacmact(n)taaatatnaat	tgcaacmact(a)taaatatnaat
22693	naaatatnaat(n)acttaatgag	naaatatnaat(c)acttaatgag
34021	tggtctcaaaa(n)caaaaaaaa	tggtctcaaaa(a)caaaaaaaa
34471	tggtctcaaaa(n)caaaaaaaa	tggtctcaaaa(a)caaaaaaaa
44899	agttacagtg(n)gtatctggct	agttacagtg(a)gtatctggct
56139	acagtgctcg(n)cataataga	acagtgctcg(g)cataataga
73435	gagtggtag(n)tgacttgctg	gagtggtag(a)tgacttgctg
74766	tcacatatt(n)gggtttaaat	tcacatatt(g)gggtttaaat
94286	taacgatatt(n)aaagctcgag	taacgatatt(a)aaagctcgag
128065	gacagatgct(n)tgagctcgag	gacagatgct(c)tgagctcgag
169207	ctccatctc(n)naaaaaaaa	ctccatctc(a)naaaaaaaa
169208	cccatcttan(n)aaaaaaa	cccatcttan(a)aaaaaaa
169210	catcttana(n)aaaaaaa	catcttana(a)aaaaaaa
169214	ctannanaa(n)aaaaaaa	ctannanaa(a)aaaaaaa
169220	naaaaaaaa(n)aaagcaaatc	naaaaaaaa(a)aaagcaaatc
169276	aaatgaatatt(n)ttgtctgct	aaatgaatatt(a)ttgtctgct
191868	cccaacacac(n)aaagttaata	cccaacacac(a)aaagttaata
191870	caacacacac(n)aaagttaata	caacacacac(a)aaagttaata
198887	attctcagag(n)ttcaatccc	attctcagag(c)ttcaatccc
203149	aatggtttg(n)caatatttc	aatggtttg(a)caatatttc
207416	ctcatggaat(n).....	ctcatggaat(c).....

----- Distribution of Quality < 40 Bases -----

#	bases
1000	900
900	800
800	700
700	600
600	500

```

400
300
200
100
0
-----
5 10 15 20 25 30 35 40
Phrap Value Range

```

Version: 1.01 gxf.
Location/Qualifiers

source 1.121700

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="12"

/clone="RP11-864J10"

/complement(1087..1398)

/rpt_family="AluSx"

3261..3665

/rpt_family="MER61C"

complement(4390..4440)

/rpt_family="WIR"

4662..4839

/standard_name="11.126A/11.126B"

/db_xref="GDB:385913"

4692..4742

/rpt_family="(TG)n"

4764..4895

/rpt_family="L1MC4"

4901..5177

/rpt_family="AluJo"

5178..5214

/rpt_family="(CAAT)n"

5215..5453

/rpt_family="AluJo"

5483..6031

/rpt_family="L1MC4"

6032..6338

/rpt_family="AluSx"

6349..6445

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

Db 79818 TAATTCCTTAACAT 79804

RESULT 8

AC010187/c

LOCUS

DEFINITION

AC010187

ACCESSION

AC010187

VERSION

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 334597)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,

Barbata,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,

Douhaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgeson,A., Hognes,M., Holloway,C., Hollins,B.,

Homi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,

Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.,

Li,J., Li,Z., Lichtenberg,O., Liu,C., Liu,J., Liu,W., Louised,H.,

Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

Mareshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,B.,

Masey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,

Miner,G., Miner,Z., Mitchell,T., Mohabac,K., Morgan,M., Morris,S.,

Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,

Nguyen,N., Nickerson,E., Nwokwenkwo,S., Oguh,M., Okunolu,G.,

Ortunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

Peters,L., Pickens,R., Prims,E., Pu,L.L., Quiles,M., Ren,Y.,

Rivers,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Savery,G.,

Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,

Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,

Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,

Taney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,

Usmani,K., Vaequez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,

Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,

Williams,G., Williamson,A., Wleczek,R., Woodson,S., Worley,K.,

Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gibbs,R.

Unpublished

2 (bases 1 to 334597)

Worley,K.C.

Direct Submission

Submitted (15-SEP-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 334597)

Worley,K.C.

Direct Submission

Submitted (09-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Apr 24, 2002 this sequence version replaced gi:15667986.

----- Genome Center

Center: Baylor College of Medicine

```

Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: NHKV
Center clone name: RP11-38909
----- Summary Statistics -----
Sequencing vector: M13:
Chemistry: Dye-primer Bodipy: 77% of reads
Chemistry: Dye-terminator Big Dye: 23% of reads
Assembly program: Phrap; version: 0.990329
Consensus quality: 299089 bases at least Q40
Consensus quality: 323681 bases at least Q30
Consensus quality: 338387 bases at least Q20
Estimated insert size: 349203; sum-of-coverage estimation
Estimated insert size: 503842; agarose-fp estimation
Quality coverage: 2x in Q20 bases; agarose-fp estimation
Quality coverage: 3x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length.
NOTE: http://www.hgsc.bcm.tmc.edu/docs/genbank\_drfc\_date.html
NOTE: This sequence may represent more than one clone.
NOTE: This is a 'working draft' sequence. It currently
consists of 40 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1
2157: contig of 2157 bp in length
2158 2257: gap of unknown length
2258 3490: contig of 323 bp in length
5491 5590: gap of unknown length
5591 8466: contig of 2876 bp in length
8467 8567: gap of unknown length
8567 10661: contig of 2095 bp in length
10662 10761: gap of unknown length
10762 13041: contig of 2280 bp in length
13042 13141: gap of unknown length
13142 16011: contig of 2870 bp in length
16012 16111: gap of unknown length
16112 18931: contig of 2820 bp in length
18932 19031: gap of unknown length
19032 22675: contig of 3644 bp in length
22676 22776: gap of unknown length
22776 25400: contig of 2625 bp in length
25401 25500: gap of unknown length
25501 29388: contig of 3888 bp in length
29389 29488: gap of unknown length
29489 32970: contig of 3482 bp in length
32971 33070: gap of unknown length
33071 36207: contig of 3137 bp in length
36208 36307: gap of unknown length
36308 39878: contig of 3571 bp in length
39879 39978: gap of unknown length
39979 43312: contig of 3334 bp in length
43313 43412: gap of unknown length
43413 47414: contig of 4002 bp in length
47415 47514: gap of unknown length
47515 50144: contig of 2650 bp in length
50165 50264: gap of unknown length
50265 54914: contig of 4650 bp in length
54915 55014: gap of unknown length
55015 57456: contig of 2442 bp in length
57457 57556: gap of unknown length
57557 60864: contig of 3308 bp in length
60865 60964: gap of unknown length
60965 65163: contig of 4199 bp in length
65164 65263: gap of unknown length
70398 70498: contig of 5135 bp in length
70499 74623: contig of 4125 bp in length

```

*	74723:	gap of unknown length	*
*	74624	78535: contig of 3812 bp in length	*
*	74774	78635: gap of unknown length	*
*	78565	82428: contig of 3794 bp in length	*
*	78636	82529: gap of unknown length	*
*	82430	88169: contig of 5640 bp in length	*
*	82530	88269: gap of unknown length	*
*	88170	94292: contig of 6023 bp in length	*
*	88270	94399: gap of unknown length	*
*	94283	101748: contig of 7357 bp in length	*
*	94393	101849: gap of unknown length	*
*	101750	109260: contig of 7411 bp in length	*
*	101850	109360: gap of unknown length	*
*	109261	116775: contig of 7415 bp in length	*
*	109361	116875: gap of unknown length	*
*	116776	123551: contig of 6676 bp in length	*
*	116876	123551: gap of unknown length	*
*	123552	123651: contig of 3952 bp in length	*
*	123652	127603: gap of unknown length	*
*	127604	136510: contig of 8807 bp in length	*
*	127704	136510: gap of unknown length	*
*	136511	144218: contig of 7608 bp in length	*
*	136611	144318: gap of unknown length	*
*	144219	158471: contig of 14152 bp in length	*
*	144319	158570: gap of unknown length	*
*	158471	171113: contig of 12543 bp in length	*
*	158571	171213: gap of unknown length	*
*	171214	188563: contig of 17350 bp in length	*
*	188564	188663: gap of unknown length	*
*	188664	222888: contig of 34225 bp in length	*
*	222889	257090: gap of unknown length	*
*	257091	257190: contig of 34102 bp in length	*
*	257191	295336: gap of unknown length	*
*	295337	295436: contig of 38146 bp in length	*
*	295437	334597: gap of unknown length	*
*	334597	39161 bp in length	*

FEATURES	Location/Qualifiers	
SOURCE	1. .334597	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="12"	
	/clone="RP11-38909"	
BASE COUNT	90733 a 73766 c 75238 g 90890 t	39770
ORIGIN		others
Query Match	20.2%	Score 373.4; DB 2;
Best Local Similarity	99.7%	Pred. No. 6e-90; 1;
Matches 374; Conservative	0; Mismatches	Indels 0; Gaps 0;

Qy	1479	TTGAGTAAATGAGCATTTGGAAAGCAGAGAGGCTTTCTTGGGAGTGCAGAAATTC	1538
Db	48112	TTTACTAATATGACATTTGGAAACGAGAGAGGCTTTCTTGGAGTGCAGAAATTC	480833
Qy	1539	ATCTATTAGTGCCCTTTTGGCAGCAGGAGCTCAGATACCCCATGAGTCTTAATGAG	1598
Db	48082	ATCTATTAGTGCCCTTTTGGCAGCAGGAGCTCAGATACCCCATGAGTCTTAATGAG	480232
Qy	1559	ATCATCTAATTTCTCGCTTACTTAACTGGGAATTTTAAAGTTGGGCTTAAATGTTTC	1658
Db	48022	ATCATCTAATTTCTCGCTTACTTAACTGGGAATTTTAAAGTTGGGCTTAAATGTTTC	479663
Qy	1659	CATGGTGCGCAGCAACATGCGAGAGTGACTGAACACAACTTCATTCTGACTGACATTC	1718
Db	47962	CATGGTGCGCAGCAACATGCGAGAGTGACTGAACACAACTTCATTCTGACTGACATTC	479033
Qy	1719	TTTCTGCACTGCGGTGCGACCAACAATATCCGGTCTTTGTGCTGTCTCTTCAGATGAT	1778
Db	47902	TTTCTGCACTGCGGTGCGACCAACAATATCCGGTCTTTGTGCTGTCTCTTCAGATGAT	478433
Qy	1779	GATTGTGAAGCTCTGTGTGATTTCTTAAAGAAAGTAATTAAGATTTTAAAG	1838
Db	47842	GATTGTGAAGCTCTGTGTGATTTCTTAAAGAAAGTAATTAAGATTTTAAAG	477833

QY 1839 TAATCTTAAACAT 1853
|||||
Db 47782 TAATCTTAAACAT 47768

RESULT 9
HSXS153
LOCUS HSXS153 400 bp mRNA linear PRI 10-APR-1997
DEFINITION H.sapiens (xsl153) mRNA, 400bp.
ACCESSION Z36805
VERSION Z36805.1 GI:533924
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 400)
REFERENCE
AUTHORS Mueller-Pillasch,F., Gress,T., Lehnach,H. and Adler,G.
TITLE Differential gene expression in pancreatic cancer. Use of an
automated approach for the large scale isolation and
characterization of cDNA clones containing differentially expressed
sequences
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 400)
AUTHORS Gress,T.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-1994) Gress T., University of Ulm, Department of
Internal Medicine I, Robert Koch Str.8, 89081 Ulm, Germany, 89081
3 (bases 1 to 400)
REFERENCE
AUTHORS Gress,T.M., Muller-Pillasch,F., Geng,M., Zimmerhackl,F.,
Zehner,G., Fries,H., Buchler,M., Adler,G. and Lehnach,H.
TITLE A pancreatic cancer-specific expression profile
JOURNAL Oncogene 13 (8), 1819-1830 (1996)
MEDLINE 97050791
PUBMED 8895530

FEATURES
source
1.400 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="xsl153"
/cell_line="PATU 8988s"
/cell_type="ductal adenocarcinoma"
/issue_type="pancreas"
/clone_lib="xs CDNA lib, pancreatic cancer cell line PATU
8988s"
/dev_stage="adult"
misc_feature
1..362
/product="56 kDa protein"
/note="95% homology to human 56 kDa protein, GenEMBL-locus
gb_pr:hum56kdap residues 419-797"
BASE COUNT 113 a 78 c 90 g 115 t 4 others
ORIGIN

Query Match 18.2%; Score 337.6; DB 9; Length 400;
Best Local Similarity 94.8%; Pred. No. 1.3e-80;
Matches 379; Conservative 0; Mismatches 18; Indels 3; Gaps 3;

QY 419 TTGGTAATCTCTTGGGCTTACGCTTACCGGAGTATGATCAAGATCCTTACCTTA 478
|||||
Db 1 TTGGTAATCTCTTGGGCTTACGCTTACCGGAGTATGATCAAGATCCTTACCTTA 60

QY 479 CTCGAAAGATCAGAACATATGAGAGTATTTCTTGATTAGCCGAGATATC 538
|||||
Db 61 CTCGAAAGATCAGAACATATGAGAGTATTTCTTGATTAGCCGAGATATC 120

QY 539 TTATAGTTTGTGGCCGAGCTGAACAGACCGAGTCAATTAGAGGTGCATGTTATATC 598
|||||
Db 121 TTATAGGTGTGGCCGAGCTGAACAGACCGAGTCAATTAGAGGTGCATGTTATATC 180

QY 599 AAGAAAGAGCTCTTTTATGATACCATATATCTTGTCTGCATATCCTTGAGTG 658
|||||
Db 181 AAGAAAGAGCTCTTTTATGATACCATATATCTTGTCTGCATATCCTTGAGTG 240

QY 659 TGAATGCTGTAAT-TTGATCTAGCCGATGATTTACTGGAATTAATTAATCTGCTGA 717
|||||
Db 241 TGAATGCTGTAATTTGATGATCTAGCCGATGATTTACTGGAATTAATTAATCTGCTGA 300

QY 718 GGAACAT-GACCCCTTTTGAAGTGGGACCTTGATATGTCGACCTTTAGAGCC 776
|||||
Db 301 GGAACATCCACCCCTTTTGAAGTGGGACCCGATATGTCGACCTTTAGAGNC 360

QY 777 AGTCTTCACACTC-GGAAGTAACCTTCAAAAAAGAA 815
|||||
Db 361 AGTCTTCACACTCGGAAGTAACTNCCAAACAGCAGAA 400

RESULT 10
AC109397
LOCUS AC109397 112300 bp DNA linear HTG 13-JUL-2002
DEFINITION Rattus norvegicus clone CH230-261M6, *** SEQUENCING IN PROGRESS
AC109397
AC109397.3 GI:21465200
HTG: HTGS PHASR1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 112300)
REFERENCE
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaralunge,H.C., Aré,J.R., Ayele,M., Banks,T.,
Barbarta,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowe,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Deleavy,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homes,F., Howard,S., Huber,J., Hulky,S., Hune,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolyet,S., Joudah,S.,
Karlssohn,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,F.,
Massey,E., Maunhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newsham,N., Nguyen,N., Nguyen,N.,
Onguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,D., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojts,A., Rojibokan,T., Rolfe,M., Ruiz,S., Savery,G.,
Seherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Taneey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Tumani,K., Vaquer,L., Vera,V., Villalón,D., Vinton,R., Wang,O.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 112300)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2002) Human Genome Sequencing Center, Department

REFERENCE
AUTHORS
TITLE
JOURNAL

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 112300)

Worley, K.C.

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Jun 19, 2002 this sequence version replaced gi:18846849.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GPZA

Center clone name: CH230-261W6

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 67877 bases at least Q40

Consensus quality: 71488 bases at least Q30

Consensus quality: 74468 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

* (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 52 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1392: contig of 1392 bp in length
* 1393 1492: gap of unknown length
* 1493 2616: contig of 1124 bp in length
* 2617 2716: gap of unknown length
* 2717 4137: contig of 1421 bp in length
* 4138 4237: gap of unknown length
* 4238 5861: contig of 1624 bp in length
* 5862 7254: gap of unknown length
* 7255 7354: contig of 1293 bp in length
* 7355 8434: gap of unknown length
* 8435 8534: gap of unknown length
* 8535 9591: contig of 1057 bp in length
* 9592 9691: gap of unknown length
* 9692 10739: contig of 1047 bp in length
* 10739 10838: gap of unknown length
* 10839 11893: contig of 1055 bp in length
* 11894 11993: gap of unknown length
* 11994 13704: contig of 1711 bp in length
* 13705 13804: gap of unknown length
* 13806 15599: contig of 1795 bp in length
* 15600 15699: gap of unknown length
* 15700 16725: contig of 1026 bp in length
* 16726 16825: gap of unknown length
* 16826 18609: contig of 1784 bp in length
* 18610 18709: gap of unknown length
* 18710 19938: contig of 1229 bp in length
* 19939 20038: gap of unknown length
* 20039 21273: contig of 1235 bp in length
* 21274 21373: gap of unknown length
* 21374 23028: contig of 1655 bp in length
* 23029 24597: gap of unknown length
* 24598 24697: contig of 1469 bp in length
* 24698 27845: gap of unknown length
* 27846 27945: contig of 3148 bp in length
* 27946 29017: gap of unknown length
* 29018 29117: gap of unknown length

29118 31039: contig of 1922 bp in length
* 31039 31139: gap of unknown length
* 31140 32589: contig of 1450 bp in length
* 32590 32689: gap of unknown length
* 32690 35113: contig of 2424 bp in length
* 35114 35213: gap of unknown length
* 35214 37124: contig of 1911 bp in length
* 37125 37224: gap of unknown length
* 37225 39052: contig of 1828 bp in length
* 39053 39152: gap of unknown length
* 39153 40880: contig of 1728 bp in length
* 40881 40980: gap of unknown length
* 40981 42452: contig of 1472 bp in length
* 42453 42552: gap of unknown length
* 42553 44335: contig of 1783 bp in length
* 44336 44435: gap of unknown length
* 44436 45624: contig of 1189 bp in length
* 45625 45724: gap of unknown length
* 45725 48195: contig of 2471 bp in length
* 48196 48295: gap of unknown length
* 48296 49958: contig of 1663 bp in length
* 49959 50058: gap of unknown length
* 50059 52259: contig of 2201 bp in length
* 52260 52359: gap of unknown length
* 52360 53726: contig of 1367 bp in length
* 53727 53826: gap of unknown length
* 53827 55827: contig of 1956 bp in length
* 55828 55882: gap of unknown length
* 55883 57544: contig of 1662 bp in length
* 57545 57644: gap of unknown length
* 57645 59237: contig of 1593 bp in length
* 59238 59337: gap of unknown length
* 59338 61091: contig of 1754 bp in length
* 61092 61191: gap of unknown length
* 61192 62543: contig of 1352 bp in length
* 62544 62643: gap of unknown length
* 62644 65077: contig of 2434 bp in length
* 65078 65177: gap of unknown length
* 65178 67369: contig of 2192 bp in length
* 67370 67469: gap of unknown length
* 67470 69466: contig of 1997 bp in length
* 69467 69566: gap of unknown length
* 69567 72140: contig of 2574 bp in length
* 72141 72240: gap of unknown length
* 72241 74313: contig of 2073 bp in length
* 74314 74413: gap of unknown length
* 74414 77865: contig of 3452 bp in length
* 77866 77966: gap of unknown length
* 77967 81535: contig of 3570 bp in length
* 81536 81635: gap of unknown length
* 81636 85369: contig of 3734 bp in length
* 85370 85469: gap of unknown length
* 85470 88583: contig of 3114 bp in length
* 88584 88683: gap of unknown length
* 88684 92585: contig of 3902 bp in length
* 92586 92685: gap of unknown length
* 92686 96502: contig of 3817 bp in length
* 96503 96602: gap of unknown length
* 96603 100074: contig of 3472 bp in length
* 100075 100174: gap of unknown length
* 100175 102758: contig of 2584 bp in length
* 102759 102858: gap of unknown length
* 102859 108435: contig of 5577 bp in length
* 108436 108535: gap of unknown length
* 108536 112300: contig of 3765 bp in length.

Location/Qualifiers

1. .112300

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

FEATURES

source

Query Match 15.4%; Score 285; DB 2; Length 112300;
Best Local Similarity 77.6%; Pred. No. 6,1e-66;
Matches 384; Conservative 0; Mismatches 105; Indels 6; Gaps 3;

QY 1067 ATGACTGCCAAGTCCAGATGAAAGCCATCGAATGTGCGATTGAGTGGCGAGATAGAGA 1126
 DB 40374 ATGACTGCACTGAAACACATCCAAATAATACCTCATGAGCACTTGAGGCGAGATCAAA 40433
 QY 1127 GATGACTTGGATCACTTTTCACTTGTATTTCTTGGCCAGTACAGATGAGCGCTTTG 1186
 DB 40434 AAGTGAACGGAACCATTTCTCACCATGATCATATCTTGGCCAGTACAGATGATGTTTTG 40493
 QY 1187 TATATAATTTGATGACGTTGAGATTAAGCCAAATTTTATCACTTAATGACACAAATGATG 1246
 DB 40494 TATATATCTGGATACACGCTTCAGAGAAAGTTGATTTTAACTTAATGACACAGACTATG 40553
 QY 1247 AAATCTGTGCTTGTATCTTAGACGCAAAATCAAGGCGTGTCTGAGTCTGCTCACTG 1306
 DB 40554 AAATGCTGTGCTTGTATCTTAGACGATAGATCAAAAGCGCTGCTGTGACTG--TCCACAG 40611
 QY 1307 ACAAAATACGGAAGATGTGGACATCTTAGAGATTAAGCCAAAGTCTAGTTCATTCTAGGG 1366
 DB 40612 ACAATTTTGGAAAGATCTGGATATCTTAGAGAAAGCCAAAGTCTCATTCACCTTAGAG 40671
 QY 1367 ACATGAATATGGAGAGTCTCTTCTGTTCTTCATGTTGCCGATTTGSCATTTATTATG 1426
 DB 40672 ACATTTAAATGAAGATCTCTT--TTCAATCAAGCTTCTGATTTGTCAATTTGTTATG 40728
 QY 1427 CCTTTGAGGTCAAAAAGAGGCTTCGGTCTGGATATAGACAGTCTCTTCACTAA 1486
 DB 40729 CCTTTGAGAGACAGAAAGAGGCTTTGGTTGGGAT--TAAGACAGTCTCTTCACTAA 40787
 QY 1487 ATGAACATTTGGAAGACGAGAGAGCTTGTCTTGGAGAGTCAAGAAATTCATCTATTA 1546
 DB 40788 ATGAATGTTTGGAGACATGAGAGACTGATTTATGGCAACAGCATTTTGGGTAATTA 40847
 QY 1547 GTGGCCCTTTGGCA 1561
 DB 40848 GTGGCCCTTAGCCA 40862

RESULT 11
 G06393 260 bp DNA linear STS 19-OCT-1995
 LOCUS human STS WI-7102, sequence tagged site.
 DEFINITION G06393
 ACCESSION G06393.1 GI:859638
 VERSION STS: STS sequence; primer: sequence tagged site.
 KEYWORDS Homo sapiens STS derived from sequences in dbEST and the Unigene collection.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 260)
 Hudson, T.
 Whitehead Institute/MIT Center for Genome Research; Physically Mapped ESTs
 Unpublished (1995)
 JOURNAL COMMENT
 CONTACT: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu
 PRIMER A: ATTTCCTGCTTACCTTAACCTGAG
 PRIMER B: CTTTATTGAAGAAATGCAACAGAGC
 STS size: 207
 PCR Profile:
 Presoak:
 Denaturation:
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35

Thermal Cycler:
 Protocol:
 Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul
 Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

Prepared with primer pairs derived from L07758 -- Unigene.

FEATURES
 source
 1..260
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="729 D 6; 744 G 5; 909 G 6; 925 G 1; 948 F 3; 959 A 1; 720 G 10; 786 F 12; 969 C 12"
 14..320
 14..36
 complement(196..220)
 primer_bind
 primer_bind
 BASE COUNT 79 a 49 c 48 g 84 t
 ORIGIN

Query Match 14.0%; Score 260; DB 11; Length 260;
 Best Local Similarity 100.0%; Pred. No. 1.7e-59;
 Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1594 TGAAGATCATCTAATTTCTGCTTACCTTAACGGAATTTTAAAAAGTGGCCTAAAA 1653
 DB 1 TGAAGATCATCTAATTTCTGCTTACCTTAACGGAATTTTAAAAAGTGGCCTAAAA 60
 QY 1654 TGTTCATGCGTGCGACGACCAACATGAGAGTGAAGTGAACACAAATTCATTCTGACTGAC 1713
 DB 61 TGTTCATGCGTGCGACGACCAACATGAGAGTGAAGTGAACACAAATTCATTCTGACTGAC 120
 QY 1714 ATTCCCTTTCTGCAACGCGGTGGCGACCAAAATACCGGCTTTGGCTTCTTCAAG 1773
 DB 121 ATTCCCTTTCTGCAACGCGGTGGCGACCAAAATACCGGCTTTGGCTTCTTCAAG 180
 QY 1774 TGGATGTTTGAAGCTCTGTTGTCATTTCTTAATAAAGATTAATAAAGATTTTAA 1833
 DB 181 TGGATGTTTGAAGCTCTGTTGTCATTTCTTAATAAAGATTAATAAAGATTTTAA 240
 QY 1834 AAAAGTAATTCCTTAACAT 1853
 DB 241 AAAAGTAATTCCTTAACAT 260

RESULT 12
 HSXS152 250 bp mRNA linear PRI 10-APR-1997
 LOCUS HSXS152
 DEFINITION H.sapiens (xs152) mRNA, 250bp.
 ACCESSION Z36804
 VERSION Z36804.1 GI:533911
 KEYWORDS
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 250)
 Mueller-Pillasch, F., Gress, T., Lehnach, H. and Adler, G.
 Differential gene expression in pancreatic cancer. Use of an automated approach for the large scale isolation and characterisation of cDNA clones containing differentially expressed sequences
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 250)
 AUTHORS Gress, T.
 TITLE Direct Submission

JOURNAL Submitted (16-AUG-1994) Gress T., University of Ulm, Department of Internal Medicine I, Robert Koch Str.8, 89081 Ulm, Germany, 89081

REFERENCE 3 (bases 1 to 250)

AUTHORS Gress,T.M., Muller-Pillasch F., Geng M., Zimmerhackl,F., Zehner,G., Friese,H., Buchler,M., Adler,G. and Lennrach,H.

TITLE A pancreatic cancer-specific expression profile

JOURNAL Oncogene 13 (8), 1819-1830 (1996)

MEDLINE 97050791

PUBMED 8895530

FEATURES

source Location/Qualifiers

1..250

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="x152"

/cell_line="PATU 8988"

/cell_type="ductal adenocarcinoma"

/tissue_type="pancreas"

/clone_id="xs cdna 11b, pancreatic cancer cell line PATU 8988"

/dev_stage="adult"

misc_feature 1..250

/product="56 kDa protein"

/note="98% homology to human 56 kDa protein, GENEMBL-locus gb:U056kDapr residues 419-669"

BASE COUNT 70 a 45 c 56 g 75 t 4 others

ORIGIN

Query Match 13.1%; Score 242.8; DB 9; Length 250;

Best Local Similarity 97.6%; Pred. No. 8e-55; Indels 0; Gaps 0;

Matches 244; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 418 CTTGTGTAATCTCTCTGGGCTCTTACGGCTTACGGAGTAATGATCAAGATCTTACGTT 477

Db 1 CTTGTGTAATCTCTCTGGGCTCTTACGGCTTACGGAGTAATGATCAAGATCTTACGTT 60

QY 478 ACTGTGAAGATACAGAACAAATATGACGTGAAGATTTCTGATTAAAGCCAGTATAT 537

Db 61 ACTGTGAAGATACAGAACAAATATGACGTGAAGATTTCTGATTAAAGCCAGTATAT 120

QY 538 CTTATAGTTTGGCCGAGCTGAAGAGACCAAGTCATTTAGAGAGTCATGTTATAT 597

Db 121 CTTATAGTTTGGCCGAGCTGAAGAGACCAAGTCATTTAGAGAGTCATGTTATAT 180

QY 598 CAGAAGAAGACTCTTTTATGTACACCATGATATCTTGTCTGATATCTCTGAGT 657

Db 181 CAGAAGAAGACTCTTTTATGTACACCATGATATCTTGTCTGATATCTCTGAGT 240

QY 658 GTGGAATGCG 667

Db 241 GTGGAATGCG 250

RESULT 13

AC010187 334597 bp DNA linear HTG 09-MAY-2002

LOCUS Homo sapiens chromosome 12 clone RP11-38909, WORKING DRAFT

DEFINITION

SEQUENCE 40 unordered pieces.

AC010187 AC010187.14 GI:20279317

VERSION

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 334597)

Murphy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alabrooks,S.L., Amaralung,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Burch,P., Burke,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Davy-Carroll,L., Dedrich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Dulaney,K.R., Draper,H., Dugan-Rocha,S., Durbin,K.D., Fairhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Faller,T., Ferraguto,D., Flagg,N.C., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabac,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguni,M., Okunnu,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojia,A., Rojibokan,I., Rolle,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tameris,A., Tameris,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Teitrod,B., Thomas,N., Thomas,S., Umami,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczky,R., Woodson,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

REFERENCE

TITLE Direct Submission

JOURNAL Unpublished

AUTHORS Worley,K.C.

TITLE

JOURNAL 2 (bases 1 to 334597)

REFERENCE

TITLE Direct Submission

JOURNAL Submitted (15-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 334597)

Worley,K.C.

Direct Submission

Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Apr 24, 2002 this sequence version replaced gi:15667986.

COMMENT

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: HMKV

Center clone name: RP11-38909

Summary Statistics

Sequencing vector: M13;

Chemistry: Dye-Primer Bodipy: 77% of reads

Chemistry: Dye-terminator Big Dye: 23% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 298069 bases at least Q40

Consensus quality: 323681 bases at least Q30

Consensus quality: 338387 bases at least Q20

Estimated insert size: 349203; sum-of-contigs estimation

Estimated insert size: 503842; agarose-fp estimation

Quality coverage: 2x in Q20 bases; agarose-fp estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently consists of 40 contigs. The true order of the pieces

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2003, 19:15:45 ; Search time 865 Seconds
(without alignments)
4824.223 Million cell updates/sec

Title: US-09-315-355a-47

Perfect score: 1853
Sequence: 1 gatccctgagcgtcgtgcag.....aaagtaattcctaacaac 1853

Scoring table: IDENTITY NUC
Gapop 10_0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_101002.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1771	95.6	1881	23	AAST70866
2	1450	78.3	2584	24	ABO54496
3	517	27.9	626	21	AAO4021
4	239.8	12.9	596	24	ABO60025
5	188.4	10.2	17197	22	AAK79870
6	150	8.1	150	24	ABN96228
7	122.8	6.6	381	21	AAH31017
8	119.8	6.5	796	24	ABO89158
9	108.8	5.9	1965	23	ABL1659

10	108.8	5.9	3965	23	ABL11658	Drosophila melanog
11	108.8	5.9	4626	23	ABL13260	Drosophila melanog
12	106	5.7	1566	21	AAO51036	Arabidopsis thaliana
13	104	5.6	553	20	AAO21042	Polynucleotide seq
14	74	4.0	261	23	AAST70865	DNA encoding novel
15	69.6	3.8	211	23	ABV17865	Human prostate exp
16	69.6	3.8	335	23	ABV47656	Human prostate exp
17	62.8	2.8	766	22	ABA49279	Human breast cell
18	52.8	2.8	766	22	ABA67193	Human foetal liver
19	52.8	2.8	766	22	ABA34288	Probe #12754 for g
20	52.8	2.8	766	22	AA122114	Probe #12047 for g
21	52.8	2.8	766	22	AA147409	Probe #16095 used t
22	52.8	2.8	766	22	AA107812	Probe #7803 used t
23	52.8	2.8	1944	22	ABA44123	Human breast cell
24	52.8	2.8	1944	22	ABA54575	Human foetal liver
25	52.8	2.8	1944	22	ABA24359	Probe #2825 for ge
26	52.8	2.8	1944	22	AA112874	Probe #2807 for ge
27	52.8	2.8	1944	22	AA134231	Probe #2917 used t
28	52.8	2.8	1944	22	AA102791	Probe #2782 used t
29	49.2	2.7	489	22	AAH06751	Human CDNA clone (
30	49	2.6	1959	22	ABA55872	Human breast cell
31	49	2.6	1959	22	ABA55872	Human foetal liver
32	49	2.6	1959	22	ABA25546	Human brain expres
33	49	2.6	1959	22	AAK04090	Human bone marrow
34	49	2.6	1959	22	AAK29571	Probe #4219 for ge
35	49	2.6	1959	22	AA114146	Probe #4217 used t
36	49	2.6	1959	22	AA135531	Probe #3985 used t
37	49	2.6	1959	22	AA103994	Human genome-deriv
38	49	2.6	1959	22	ABO4127	Human CDNA sequenc
39	48.4	2.6	1893	22	AAH14157	Human V3 loop HIV
40	48.4	2.6	2518	20	AAV71145	Human nucleolin DN
41	48.4	2.6	2518	22	AAO92553	Human prostate exp
42	48.4	2.6	3201	23	ABV23046	Human prostate exp
43	48.4	2.6	3201	23	ABV23046	Human prostate exp
44	47.8	2.6	583	22	ABA50512	Human breast cell
45	47.8	2.6	583	22	ABA68463	Human foetal liver

ALIGNMENTS

RESULT 1	AAST70866	standard; cDNA; 1881 BP.
XX	AAST70866;	
AC	AAST70866;	
XX		
DT	13-FEB-2002	(first entry)
XX		
DE	DNA encoding novel human diagnostic protein #6670.	
XX		
KM	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
XX		
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Dymanac RT, Liu C, Tang YR;	
XX		
DR	WPI; 2001-639352/73.	
XX		
DR	P-PSDB; ABO6679.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 6670; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for creating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1881 BP; 530 A; 376 C; 473 G; 502 T; 0 other:
Query Match 95.6%; Score 1771; DB 23; Length 1881;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1843; Conservative 0; Mismatches 10; Indels 6; Gaps 6;
QY 1 GATCCCTGAGCGCTGCGACGAGTGGCTGCTGCTCCCTTATGCACTT-GGTTCT 59
DB 23 GATCCCTGAGCGCTGCGACGAGTGGCTGCTGCTCCCTTATGCACTTGGTTCT 82
QY 60 AGCGGACACGCGCTTGAATTGAGACCATGAAACGCGCGCCAGGAGACGTGGTGGC 119
DB 83 AGCGGACACGCGCTTGAATTGAGACCATGAAACGCGCGCCAGGAGACGTGGTGGC 142
QY 120 CTGGGCTCCGCTGCGCGCTGCGCAAGAGACACACAGACAGGTAGAGCTGATTAAGAA 179
DB 143 CTGGGCTCCGCTGCGCGCTGCGCAAGAGACACACAGACAGGTAGAGCTGATTAAGAA 202
QY 180 AGTAAACGCTCTATCTGCTGAGGCAAGAGAAATTGCAAGAAAGTGTGGCACTGA 239
DB 203 AGTAAACGCTCTATCTGCTGAGGCAAGAGAAATTGCAAGAAAGTGTGGCACTGA 262
QY 240 TGAAGAGAGACAGGACGCTTTCAGAAAGATGGCA-TGCAGAGTGCACGCGACCCAGGAC 298
DB 263 TGAAGAGAGACAGGACGCTTTCAGAAAGATGGCA-TGCAGAGTGCACGCGACCCAGGAC 322
QY 299 G-CCCAAGAGAGCCCTGGA-GGATGTGACCCAGAGGA-TGACAGAGCGCTTGAATG 355
DB 323 GCCCAAGAGAGCCCTGGAAGGATGTGACCCAGAGGATGACAGAGCGCTTGAATG 382
QY 356 ATGAGCTGCTGAGTACGACTTATTAATATGATGAGAGAGTACCCAGATGCTGAGA 415
DB 383 ATGAGCTGCTGAGTACGACTTATTAATATGATGAGAGAGTACCCAGATGCTGAGA 442
QY 416 CTCTTGG-TGAATCTCTCTGGGCTTACCGCTTACCGGAGTAATGATCAATATCTTAC 474
DB 443 CTCTTGGTGAATCTCTCTGGGCTTACCGCTTACCGGAGTAATGATCAATATCTTAC 502
QY 475 GTTACTCTGAAGAATACAGAAACAATATGAAACGTGAAGATTTCTTATTAAGCCAGTAT 534
DB 503 GTTACTCTGAAGAATACAGAAACAATATGAAACGTGAAGATTTCTTATTAAGCCAGTAT 562
QY 535 AATCTTATAGTTTGTGGCGAGAGCTGAACAGACAGTGCATTTTATGAGGTGATTTAT 594
DB 563 AATCTTATAGTTTGTGGCGAGAGCTGAACAGACAGTGCATTTTATGAGGTGATTTAT 622

QY 595 AATCAAGAAGAAGACTTTTATATGACACCATGATATCTTGTGATATCTCTG 654
DB 623 AATCAAGAAGAAGACTTTTATATGATACATGATATCTTGTGATATCTCTG 682
QY 655 AGTGTGAATGCTGAATTTTGAATCTTACCCAGATGATTTCTGCAATTTACTGCT 714
DB 683 AGTGTGAATGCTGAATTTTGAATCTTACCCAGATGATTTCTGCAATTTACTGCT 742
QY 715 GTAGAAACATGACCCCTGATTAAGTGGGACCTTGAATATAGTGAATCTTTAGAG 774
DB 743 GTAGAAACATGACCCCTGATTAAGTGGGACCTTGAATATAGTGAATCTTTAGAG 802
QY 775 CCACTCTTCACTCGAAGTAACTTTCAAAAAAGAAAGAAAGAAAGAGAGT 834
DB 803 CCACTCTTCACTCGAAGTAACTTTCAAAAAAGAAAGAAAGAAAGAGAGT 862
QY 835 TCTTCAGAGAAAGGACATACCGATGCTGCTCTTGAACCTTTCATGATTAATCAGA 894
DB 863 TCTTCAGAGAAAGGACATACCGATGCTGCTCTTGAACCTTTCATGATTAATCAGA 922
QY 895 AATGTTTAGAAGTGCATTCAGCTGACACCTGAATTCCTGGGATATGCTTGGGG 954
DB 923 AATGTTTAGAAGTGCATTCAGCTGACACCTGAATTCCTGGGATATGCTTGGGG 982
QY 955 AAACAGCAGCTAGCTGCTGCTGACACACAGACAGAGTCCAAACAGTGCATTTTATCCA 1014
DB 983 AAACAGCAGCTAGCTGCTGCTGACACACAGAGTCCAAACAGTGCATTTTATCCA 1042
QY 1015 TTGGAAGCAGACACTGATTTTCTGCTCATATGATTAAGTGAAGTCTTGTATGACTGC 1074
DB 1043 TTGGAAGCAGACACTGATTTTCTGCTCATATGATTAAGTGAAGTCTTGTATGACTGC 1102
QY 1075 CGAAGTCAGATGAAGGACATGAGATGAGGATTCAGTGGGAGATTAAGAGAGTACT 1124
DB 1103 CGAAGTCAGATGAAGGACATGAGATGAGGATTCAGTGGGAGATTAAGAGAGTACT 1162
QY 1135 TGAATCACTTTTCACTTGTCTTCTTGGCAGTACAGATGAGGCTTTGTATATAT 1194
DB 1163 TGAATCACTTTTCACTTGTCTTCTTGGCAGTACAGATGAGGCTTTGTATATAT 1222
QY 1195 TTGATGACGCTTCAATTAAGCCAAATTTTACCTTAATGACACATGATGAATCTCT 1254
DB 1223 TTGATGACGCTTCAATTAAGCCAAATTTTACCTTAATGACACATGATGAATCTCT 1282
QY 1255 GGTCTTGAATCTTACAGTCAATCAAGGCTGCTCGGATGCTTCAAGTCAAGTCAATAC 1314
DB 1283 GGTCTTGAATCTTACAGTCAATCAAGGCTGCTCGGATGCTTCAAGTCAAGTCAATAC 1342
QY 1315 GTGAAGATCTGGACATCTTGAAGATAGGCAAGTCAATGTTCAATTTAGAGGACATGAA 1374
DB 1343 GTGAAGATCTGGACATCTTGAAGATAGGCAAGTCAATGTTCAATTTAGAGGACATGAA 1402
QY 1435 GGTCAAAAAAGAGGCTTGGGCTGAGATTAAGCAAGTCTTCAAGTCAATTAAGAGCA 1494
DB 1463 GGTCAAAAAAGAGGCTTGGGCTGAGATTAAGCAAGTCTTCAAGTCAATTAAGAGCA 1522
QY 1495 TTTGGAAGACAGAGAGGCTTGTCTTGGAGTCAAGAAATTCATCTAATAGGAGCCT 1554
DB 1523 TTTGGAAGACAGAGAGGCTTGTCTTGGAGTCAAGAAATTCATCTAATAGGAGCCT 1582
QY 1555 TTTGGAAGACAGAGCTCAGATACACCAATGAGTCTTAAATGAGATCATTAATTTCTG 1614
DB 1583 TTTGGAAGACAGAGCTCAGATACACCAATGAGTCTTAAATGAGATCATTAATTTCTG 1642
QY 1615 CTTACCTTAATCTGGGAATTTTAAAGTTGGCTTAAAGTTTCAAGCGCTGGAGCAAC 1674
DB 1643 CTTACCTTAATCTGGGAATTTTAAAGTTGGCTTAAAGTTTCAAGCGCTGGAGCAAC 1702

QY 1675 CATGACAGTACTGAAACAATTCATTCTGACCTGCTCTTCTGCAACCTGGGT 1734
 DB 1703 CATGACAGTACTGAAACAATTCATTCTGACCTGCTCTTCTGCAACCTGGGT 1762
 QY 1735 GGCACCAACAATTCCTGGTCTTGTGCTCTTCTGACATGATGTTTGAAGGCTCTT 1794
 DB 1763 GGCACCAACAATTCCTGGTCTTGTGCTCTTCTGACATGATGTTTGAAGGCTCTT 1822
 QY 1795 GTTGCAATTCCTTAAAAAGAGTATAAAAAAGATTTTAAAAAGTAACTTCAACAT 1853
 DB 1823 GTTGCAATTCCTTAAAAAGAGTATAAAAAAGATTTTAAAAAGTAACTTCAACAT 1881

RESULT 2
 AB054496
 ID AB054496 standard; cDNA; 2584 BP.
 XX AB054496;
 AC
 XX
 XX 22-AUG-2002 (first entry)
 DT
 XX
 XX Human ovarian antigen HCOQH27 cDNA, SEQ ID NO:376.
 DE
 XX
 XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KM ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KM infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KM PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 KM inflammatory condition; immune disorder; blood disorder;
 KM cardiovascular disorder; respiratory disorder; neurological disorder;
 KM gastrointestinal disorder; urinary system disorder; drug screening;
 KM gene therapy; chromosome mapping; forensic analysis;
 KM antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KM antiinflammatory; gynaecological; reproductive; chromosome 12;
 KM gene; ss.

OS Homo sapiens.
 XX
 XX WO200200677-A1.
 PN
 XX
 XX 03-JAN-2002.
 PD
 XX
 XX 07-JUN-2001; 2001WO-US18569.
 PF
 XX
 XX 07-JUN-2000; 2000US-209467P.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Birse CE, Rosen CA;
 PI
 XX
 XX WPI: 2002-147878/19.
 DR
 XX
 XX P-PSDB; ABP41419.
 XX
 XX
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 PT
 XX
 XX Claim 1, SEQ ID No 376; 2922bp; English.
 PS
 XX
 XX The invention relates to 2175 novel human ovarian antiens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and

CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents cDNA encoding a human ovarian antigen of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 2584 BP; 804 A; 454 C; 529 G; 796 T; 1 other;
 SQ

Query Match 78.3%; Score 1450; DB 24; Length 2584;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 CAGATGCTGAGACTCTTGTGTAATCTCTTGGGTCTTACGCTTACGGAGTATGATC 463
 DB 129 CAGATGCTGAGACTCTTGTGTAATCTCTTGGGTCTTACGCTTACGGAGTATGATC 188
 QY 464 AAGATCCTTACCTTACTCTGAAAGATACAGAAATATGAAAGCTTGAAGATTTCTGATTA 523
 DB 189 AAGATCCTTACCTTACTCTGAAAGATACAGAAATATGAAAGCTTGAAGATTTCTGATTA 248
 QY 524 AGCCAGTGTATCTTATATGTTTGTGCGCAGCTGAAACGAGCCAGTGAATTTAGAG 583
 DB 249 AGCCAGTGTATCTTATATGTTTGTGCGCAGCTGAAACGAGCCAGTGAATTTAGAG 308
 QY 584 TGCATGTTTAAATCAAGAAAGAACTCTTTATGTACACATATATCTTGTCTG 643
 DB 309 TGCATGTTTAAATCAAGAAAGAACTCTTTATGTACACATATATCTTGTCTG 368
 QY 644 CATATCCTCTGAGTGTGGAATGGCTGAATTTTATCTTACCCAGATGATTTACAGGA 703
 DB 369 CATATCCTCTGAGTGTGGAATGGCTGAATTTTATCTTACCCAGATGATTTACAGGA 428
 QY 704 ATTACATTTGCTGTAGGAAACATGACCCCTGTATTTGAAGTGTGGACCTTGATTAAGTG 763
 DB 429 ATTACATTTGCTGTAGGAAACATGACCCCTGTATTTGAAGTGTGGACCTTGATTAAGTG 488
 QY 764 ACTCTTTAGACCAAGTCTTACACTCGGAAATTAACCTTTCAAAAAAGAAAAAGAAAG 823
 DB 489 ACTCTTTAGACCAAGTCTTACACTCGGAAATTAACCTTTCAAAAAAGAAAAAGAAAG 548
 QY 824 GAAAGAAAGATTTCTCAGCAAGAAAGGATACCGATGTCTCTTGAACCTTTCAATGAAATA 883
 DB 549 GAAAGAAAGATTTCTCAGCAAGAAAGGATACCGATGTCTCTTGAACCTTTCAATGAAATA 608
 QY 884 AGCTAATCAGAAATGTTTGTAGCAAGTGCATCAGCTGCAACACTGTAAATCTGAGGATA 943
 DB 609 AGCTAATCAGAAATGTTTGTAGCAAGTGCATCAGCTGCAACACTGTAAATCTGAGGATA 668
 QY 944 TGTCTTGGGAAACACAGCACTAGCTGCTGTACACACAGCAAGGTCACCAACTGTC 1003
 DB 669 TGTCTTGGGAAACACAGCACTAGCTGCTGTACACACAGCAAGGTCACCAACTGTC 728
 QY 1004 AGTTTATCATTGTGAAGCAACAGCTCTGATTTCTGCTCATATGATTAAGTCAAGTGCTT 1063
 DB 729 AGTTTATCATTGTGAAGCAACAGCTCTGATTTCTGCTCATATGATTAAGTCAAGTGCTT 788
 QY 1064 TGTATGACTCCGAGAGTCCGATGAAAGCATGAAATGTGTGGCAATTCAGTGGCAGATAG 1123
 DB 789 TGTATGACTCCGAGAGTCCGATGAAAGCATGAAATGTGTGGCAATTCAGTGGCAGATAG 848
 QY 1124 AGAGAGTGAATGAAATCACTTTTCACTTTGATTTCTTGGCAGTACAGATGACGGCT 1183

Db 849 AGAGAGTACTGGAAATCATTTCACCTTGCTGTCCTTGCCAGTACAGATGCGCT 908
 Qy 1184 TTGTATATATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1243
 Db 909 TTGTATATATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 968
 Qy 1244 ATGAATCTCTGCTCTTGTATCTTACAGATCAATCAAGGCTGCTCTGCTGCTTCA 1303
 Db 969 ATGAATCTCTGCTCTTGTATCTTACAGATCAATCAAGGCTGCTCTGCTGCTTCA 1028
 Qy 1304 CTGCAATATAGTGAAGATCTGGGACATCTTGGAGATGAGCAAGTCTAGTCAATGCA 1363
 Db 1029 CTGCAATATAGTGAAGATCTGGGACATCTTGGAGATGAGCAAGTCTAGTCAATGCA 1088
 Qy 1364 GGGACATGAAATGGAGATCTCTTCTGTTCTTCAATGTTGCTGATTTGCTTATTT 1423
 Db 1089 GGGACATGAAATGGAGATCTCTTCTGTTCTTCAATGTTGCTGATTTGCTTATTT 1148
 Qy 1424 ATGCTTTGAGAGTCAAAAAGAGGCTTGGGCTTGGGATTAAGCACAGTCTTTCAG 1483
 Db 1149 ATGCTTTGAGAGTCAAAAAGAGGCTTGGGCTTGGGATTAAGCACAGTCTTTCAG 1208
 Qy 1484 TAAATGAAGCAATTTGGAGAGAGAGAGGCTTGTCTTGGAGTGCAGAAATTCATCTA 1543
 Db 1209 TAAATGAAGCAATTTGGAGAGAGAGGCTTGTCTTGGAGTGCAGAAATTCATCTA 1268
 Qy 1544 TTAGTGGCCCTTTTGGAGAGAGAGAGCTCAGATACACCATGAGCTTAAATGAGATCAT 1603
 Db 1269 TTAGTGGCCCTTTTGGAGAGAGAGAGCTCAGATACACCATGAGCTTAAATGAGATCAT 1328
 Qy 1604 CTAATTTCTGCTTACCTTAACCTGGAATTTTAAAGTTGGCTTAAATTTTCAATGC 1663
 Db 1329 CTAATTTCTGCTTACCTTAACCTGGAATTTTAAAGTTGGCTTAAATTTTCAATGC 1388
 Qy 1664 GTGGCAGACCATGACAGAGTACTGAAACACAAATTCATTTCTGACTGACATTCCTTCT 1723
 Db 1389 GTGGCAGACCATGACAGAGTACTGAAACACAAATTCATTTCTGACTGACATTCCTTCT 1448
 Qy 1724 GCAACTGGGCTGGACACCAATATCCGGTCTTGTGCTTCTTTCAGATGATGATTT 1783
 Db 1449 GCAACTGGGCTGGACACCAATATCCGGTCTTGTGCTTCTTTCAGATGATGATTT 1508
 Qy 1784 GTAAGCTCTTGTGCTTCTTAAAGAGATTAATTAAGATTTTAAAGATTAATTT 1843
 Db 1509 GTAAGCTCTTGTGCTTCTTAAAGAGATTAATTAAGATTTTAAAGATTAATTT 1568
 Qy 1844 CCTTAAACAT 1853
 Db 1569 CCTTAAACAT 1578
 RESULT 3
 AAC04021
 ID AAC04021 standard; cDNA; 626 BP.
 XX
 AC AAC04021;
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 4019.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EPI033401-A2.
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.

XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 XX
 DR P-PSDB; AN004015.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1; SEQ ID 4019; 71bp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 CC
 XX
 SQ Sequence 626 BP; 162 A; 142 C; 187 G; 135 T; 0 other.
 Query Match 27.9%; Score 517; DB 21; Length 626;
 Best Local Similarity 95.2%; Pred. No. 2,3e-118;
 Matches 555; Conservative 0; Mismatches 0; Indels 28; Gaps 1;
 Qy 1 GATCCCTGAGGCTGGGAGAGTGGCGGTGCTGCTCCCTCTTATGACGCTGATTTCTA 60
 Db 44 GATCCCTGAGGCTGGGAGAGTGGCGGTGCTGCTCCCTCTTATGACGCTGATTTCTA 103
 Qy 61 GCGTGACACGCGCTTGAATTGAGACATGAAACCGACCGCAGATGACCTGCGTGGCC 120
 Db 104 GCGTGACACGCGCTTGAATTGAGACATGAAACCGACCGCAGATGACCTGCGTGGCC 163
 Qy 121 TGGTCCGCTGCGGCGTGGCCAAAGACACACGACCAAGATGAGCTGATTAAGAA 180
 Db 164 TGGTCCGCTGCGGCGTGGCCAAAGACACACGACCAAGATGAGCTGATTAAGAA 223
 Qy 181 GTAAACCGCTCATTTGCTGAGGCAAGGAATTTGCAAGAAAGGCTGGAGAT 240
 Db 224 GTAAACCGCTCATTTGCTGAGGCAAGGAATTTGCAAGAAAGGCTGGAGAT 283
 Qy 241 GAAAGAGACAGCAGTCTTCAAGAGATGAGTGCAGATGACGACCCAGCAGC 300
 Db 284 GAAAGAGACAGCAGTCTTCAAGAGATGAGTGCAGATGACGACCCAGCAGC 343
 Qy 301 CCAAGAGACCGCTTGAAGATGAGTGCAGAGATGACAGAGCTTGAATGATGAG 360
 Db 344 CCAAGAGACCGCTTGAAGATGAGTGCAGAGATGACAGAGCTTGAATGATGAG 403
 Qy 361 CTGGCTGAGTACGACTTGAATATGATGAGGAAGTGAACCAAGTCTGAGATCTT 420
 Db 404 CTGGCTGAGTACGACTTGAATATGATGAGGAAGTGAACCAAGTCTGAGATCTT 463
 Qy 421 GGTGATCTCTCTTGGGCTTAAGGTCTACGGAGTATGATCAAGATCTTACCTTACT 480
 Db 464 GGTGATCTCTCTTGGGCTTAAGGTCTACGGAGTATGATCAAGATCTTACCTTACT 523
 Qy 481 CTGAAGATACA-----GAACAATATGAACGTGAAGA 512
 Db 524 CTGAAGATACATCAATGACATTTTTCCTCCTCAGCTTAGAACAATATGAACGTGAAGA 583
 Qy 513 TTCTTGATTAAGCCCACTGATTAATCTTATATGTTTGGCCGA 555

DB 584 TTTCTGATTAGCCGAGATATCTTATGTTTGCCGA 626

RESULT 4
AB060025
ID AB060025 standard; cDNA; 596 BP.
XX
AC AB060025;
XX
DT 02-AUG-2002 (first entry)
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:3720.
XX
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
XX genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200229086-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30732.
XX
PR 02-OCT-2000; 2000US-237271P.
XX
PA (FARB) BAYER CORP.
XX
PI Burgess C, Aetle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thiglingam A, Lewis ME;
DR WPI; 2002-426115/45.
XX
PT New isolated nucleic acid that is differentially expressed in cancer
XX tissues useful for determining the presence of colon cancer in a cell
XX or tissue type, and in antisense therapy -
XX
PS Claim 1; Fig 1; 796bp; English.

AB06306 to AB060787 represent isolated nucleic acids (I) differentially
CC expressed in cancer tissues. ABB78893 to ABB79004 represent proteins
CC encoded by the AB060776 to AB060787 nucleic acid sequences. (I) can be
CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (I) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a
CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived
CC from (I) can be used for determining the presence of a nucleic acid which
CC hybridizes to (I), and for determining the phenotype of cells in a sample
CC of cells from a patient. (I) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence or
CC state of other type of cancer, in antisense therapy, to generate
CC macroarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists.
XX
SQ

Sequence 596 BP; 179 A; 120 C; 116 G; 158 T; 23 other;

Query Match 12.9%; Score 239.8; DB 24; Length 596;
Best Local Similarity 97.2%; Pred. No. 1.7e-58;
Matches 241; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 371 AGCACTTAATAATATGATGAGGAGTGACCCAGATGCTGAGACTCTTGGAATCTC 430
DB 1 AGCACTTAATAATATGATGAGGAGTGACCCAGATGCTGAGACTCTTGGAATCTC 60
QY 431 TCTTGGCTCTTACGCTCTACGGAGTAATGATCAAGATCCTTACGTTACTTGAAGATA 490
DB 61 TCTTGGCTCTTACGCTCTACGGAGTAATGATCAAGATCCTTACGTTACTTGAAGATA 120
QY 491 CAGAAACAATATGACGTAAGATTTCTTGATTTAAGCCAGTGAATCTTATAGTTGTG 550
DB 121 CAGAAACAATATGACGTAAGATTTCTTGATTTAAGCCAGTGAATCTTATAGTTGTG 180

QY 551 GCCGACTGAACAGACGACGATTTAGAGTGATGCTTTATATCAAGAGACT 610
DB 181 GCCGACTGAACAGACGACGATTTAGAGTGATGCTTTATATCAAGAGACT 240
QY 611 CTTTATAT 618
DB 241 CTTTATAT 248

RESULT 5
AAK79870/c
ID AAK79870 standard; DNA; 17197 BP.
XX
AC AAK79870;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34682.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225470.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.


```
RESULT 6
ABN96228/c
ID ABN96228 standard; DNA; 150 BP.
XX
AC ABN96228;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #2726 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX metastatic liver tumor; cytostatic; expression profile; disease state;
XX disease progression; drug toxicity; drug efficacy; drug metabolism.
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Petes-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
XX hepatocellular carcinoma or metastatic liver tumor in a patient,
XX involves detecting the level of expression of two or more genes in a
XX liver tissue sample
XX
PS Claim 1; SEQ ID NO 2726; 298bp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumor in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 150 BP; 44 A; 33 C; 28 G; 45 T; 0 other;
XX
Query Match 8.1%; Score 150; DB 24; Length 150;
XX Best Local Similarity 100.0%; Pred. No. 5.7e-33;
XX Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 660 GGAATGGCTGAATTGATCTGACCCAGATGATCTACTGAAATTAATCTGTAGG 719
XX |||||||
DB 150 GGAATGGCTGAATTGATCTGACCCAGATGATCTACTGAAATTAATCTGTAGG 91
XX |||||||
OY 720 AAACATGACCCCTGTTATTAAGTGTGGACCTTGATATATAGTGAGCTTTAGAGCCAGT 779
XX |||||||
DB 90 AAACATGACCCCTGTTATTAAGTGTGGACCTTGATATATAGTGAGCTTTAGAGCCAGT 31
XX |||||||
OY 780 CTTGACACTCGGAAGTAACTTTCAAAAA 809
XX |||||||
DB 30 CTTGACACTCGGAAGTAACTTTCAAAAA 1
XX |||||||
```

```
RESULT 7
AAH31017
ID AAH31017 standard; cDNA; 381 BP.
XX
AC AAH31017;
XX
DT 27-JUL-2001 (first entry)
XX
DE Human colon cancer cell line Km12L4-A cDNA library derived sequence #951.
XX
KW Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
XX detection; colon cancer cell line Km12L4-A; ss.
XX
OS Homo sapiens.
XX
PN WO200018916-A2.
XX
PD 06-APR-2000.
XX
PF 23-SEP-1999; 99WO-US22226.
XX
PR 28-SEP-1998; 98US-0102161.
XX
PR 28-SEP-1998; 98US-0102180.
XX
PR 29-SEP-1998; 98US-0102380.
XX
PR 08-OCT-1998; 98US-0103815.
XX
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
XX Leshkowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;
XX
DR WPI; 2000-293155/25.
XX
PT Polynucleotide library comprising 1079 defined sequences, useful in
XX the form of an array to detect cancer or susceptibility to cancer -
XX
PS Claim 1; Page 455; 502bp; English.
XX
CC The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH3067 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides
CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive
CC interventions. Polynucleotide probes based on the disclosed sequences
CC are useful for chromosome mapping and detection of transcription levels.
CC The 1079 polynucleotide sequences were derived from a human colon cancer
CC cell line Km12L4-A cDNA library.
XX
SQ Sequence 381 BP; 67 A; 115 C; 129 G; 70 T; 0 other;
XX
Query Match 6.6%; Score 122.8; DB 21; Length 381;
XX Best Local Similarity 94.8%; Pred. No. 6.6e-25;
XX Matches 127; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
XX
OY 38 CTCCTATGACAGCTGGTCTTCTAGCGTGACAGCCCTTGACTTGAGGACATGAACCGCA 97
XX |||||||
DB 1 CTCCTATGACAGCTGGTCTTCTAGCGTGACAGCCCTTGACTTGAGGACATGAACCGCA 60
XX |||||||
OY 98 GCGCGAGGTGACGTGCGCTGGGTCCGCTGGCGGCGCTGGCCAAAGAGACACCGACA 157
XX |||||||
```

Db 61 GCCGCCAGTGACGTGGCTGGGTCCGCTCGCGGCGTGCCCAAGAGACACAGACA 120
 QY 158 AGGTAGAGCTGAGT 171
 Db 121 AGGTAGAGCTGAGT 134

RESULT 8
 ABQ89158/c
 ID ABQ89158 standard; cDNA; 796 BP.
 XX
 AC ABQ89158;

DT 27-SEP-2002 (first entry)

DE Human prostate expressed polynucleotide SEQ ID NO 414.

KW Human; prostate; cytostatic; tumour; cancer; vaccine; gene therapy;
 KW gene; ss.

OS Homo sapiens.

PN WO200255700-A2.

PD 18-JUL-2002.

PF 07-DEC-2001; 2001WO-US47349.

PR 07-DEC-2000; 2000US-254648P.

PR 13-MAR-2001; 2001US-275688P.

PA (CHIR) CHIRON CORP.
 (HYSE-) HYSEQ INC.

PI Escobedo J, Garcia PD, Kasam A, Lamson G, Drmanac R;
 PI Ckvenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D;
 PI Kita D, Garcia V, Jones WL, Stachne-Crain B, Scott EM;

XX WPI; 2002-557824/59.

PT New genes and gene products isolated from human prostate, useful for
 PT treating or diagnosing tumor or cancer (e.g. prostate cancer or breast
 PT cancer), or as vaccines for treating or preventing these diseases -
 XX
 PS Claim 1; SEQ ID NO 414; 186pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide comprising any of
 CC 1477 sequences or its fragment, degenerate variant, antisense or
 CC complement. The polynucleotides and gene products are useful for treating
 CC or diagnosing tumour or cancer (e.g. prostate cancer, breast cancer,
 CC lung cancer or medullary carcinoma) in a subject (e.g. cattle, dogs,
 CC cats, rabbits, horses or human). The polynucleotides and polypeptides are
 CC also useful as vaccines for treating or preventing these diseases. The
 CC polynucleotides are useful for gene therapy. The present sequence is that
 CC of one of a group of polynucleotides (ABQ88745-ABQ90015) disclosed
 CC electronically as sequences of the invention. However only 1271
 CC polynucleotide sequences are given, whereas 1477 polynucleotides and 91
 CC proteins are claimed.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequence.

CC Sequence 796 BP; 254 A; 161 C; 170 G; 178 T; 33 other;

QY Query Match 6.5%; Score 119.8; DB 24; Length 796;

Best Local Similarity 94.7%; Pred. No. 7.4e-24;
 Matches 124; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY .589 GTTATATCAAGAGAGACTCTTTTATGTACACCATGATATATCTTGTCTGCATAT 648

Db 307 GTTATATCAAGAGAGACTCTTTTATGTACACCATGATATATCTTGTCTGCATAT 248

QY 649 CCTGTGAGTGTGAATGGCTGAATTTGATCTTACGCCAGATGATTTACTGGAATTAC 708

Db 247 CCTGTGAGTGTGAATGGCTGAATTTGATCTTACGCCAGATGATTTACTGGAATTAG 188
 QY 709 ATTGCTGTAGG 719
 Db 187 AAAACTTAAGG 177

RESULT 9
 ABL11659
 ID ABL11659 standard; cDNA; 1965 BP.
 XX
 AC ABL11659;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 29459.

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.
 DR P-PSDB; ABB67556.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 29459; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABB5737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 1965 BP; 530 A; 486 C; 510 G; 439 T; 0 other;

QY Query Match 5.9%; Score 108.8; DB 23; Length 1965;

Best Local Similarity 48.0%; Pred. No. 1.8e-20;
 Matches 462; Conservative 0; Mismatches 467; Indels 33; Gaps 4;

QY 493 GACACATATGAACGTGAAGATTTCTTGATTAAGCCAGTGAATCTTATAGTTGGC 552

Db 425 GAAAGCTCCGAGCGCGAGAGAGAGTATCAAGCCAGCAACCTCATCTAGTGGT 484

QY 553 CGAGCTGAACGAGCAAGCTGCAATTTAGAGTGATCTTTATATCAAGAGAGACTCT 612

Db 485 CACGTTCAAGCGAGCGCGCTCCATGAGAGTGGGTTTCAACCGAGAGAGAGGCT 544

QY 613 TTTATGTACACCATGATATATCTTGTCTGCATATCTTGAAGTGGAGTGGCTGAT 672

```

Db 545 CTCTACACCCACGACGATTCTGCTGCCAAGCTTCTCTGTCATCGATGATGAAT 604
Qy 673 TTTGATCTAGCCAGATGATTTCTACTGAAATTAATTCATGCTGATGAGAAACCTCT 732
Db 605 CAGAGCGCGGCGAGCGGAAA---GGCGGCAACATGTGCCCATCGGCTGATGATCCG 661
Qy 733 GTTATTGAAGTGTGGACCTTTGATATATGTGACTCTTTAAGCCAGTCTTCAACTCGGA 792
Db 662 ATATACAGCTCTGGGATCTAGACATACAGACGCTATCGAGCCCATTTAAGCTGGT 721
Qy 793 AGTAACTTTCAAAAGAAAAGAAAAGAAAAGAAAGAGTCTCTAGACAGAAAGGCAT 852
Db 722 TCCAAAGGACCGCGAAGCAAGAACAAAGAA-----GCAATATGACAC 763
Qy 853 ACCGATGCTGTCTTGAACCTTTATGGAATAGCTAATCAAAAATGTTTATGACAGTGA 912
Db 764 AAGGAGCGCGCTGTGATCTCTTGTGAAACACCAACTTTAGACATTTGGCCAGGGG 823
Qy 913 TCAAGTGAACAACGTGATTTCTGTGGATATGTCTTTGGGAAAACGAGACGTAGCTTC 972
Db 824 TCCGTGGACCAAACTGTGATTTCTGTGGACATGAGACGAGGGCCAGCTCATACCACTT 883
Qy 973 GCGTACACACACAGCAAGGTCACAAACCTGACGTTTATTCATTTGAAGACAGACTCTG 1032
Db 884 ACCGCTTTTGGCAACAGATTCAGTGTGTAATTCATTCGCAAGGCTCAAAAGCATT 943
Qy 1033 ATTTCTGGCTCATATGATATAGTCAAGTGGCTTTTGTATGACTGCCAAGTCCAGATGA 1092
Db 944 CTTACCGGCTGTGCCGATGATAGTACGTGCACTTTCGATTTGCCGACGCTGAGGGGCTC 1003
Qy 1093 CA-----TCGAATGTGGCGATTCAGTGGCGATGAGAGAGTGAAGTGAATC 1143
Db 1004 AACTGTCGACGATTTGAGTGAAGATTTGACGATGAAATGAGAAAGTCTGTGGCATCCC 1063
Qy 1144 TTTTACCTTTGATTTCTTTGGCCAGTACAGATGAGAGCGCTTTGATATATTGATGATCA 1203
Db 1064 ACACAGACCGACTACTTCATCTGTGGGACCAACAGATGAGCACTTTCATTCGCGGACAAA 1123
Qy 1204 CGTTC---AGATAAGCCAAATTTTACACTTAATGACACAAATGATGAATCTGTGCTT 1260
Db 1124 CGTTCCTTCCGAAACCTGCTGTGTCGTAAAGGCCCAACAGAAATCTCCGCTGTG 1183
Qy 1261 GATCTTAGACGTCAAATTCAGAGGCTGTCTGTGATGCTGCTTCAAGTGAACAAATACGTA 1320
Db 1184 TGGTCAACACACAGAAAGCTTAATCTGTGACCTCCACCTCCACGAGGCGACCCCTAAG 1243
Qy 1321 ATCTGGACATCTTAGAGATAGAGCCAAAGTCTGATTTATTTAGGACATGAAATGGGA 1380
Db 1244 GTGTGAACTTTGATGACACAGAGCAAGCAAGTCTTACGAGACAGATTCACATGGGT 1303
Qy 1381 GTTCTCTTCTGTTCTTATGATGTGGCCGATTTTGGCATTTTATATGCTTTGAGGGTCAA 1440
Db 1304 CGCTTGACATGATGAGCCAGTGCCTCCGAGAGATCCCTACCTCGGCTTTCGGGGAGAG 1363
Qy 1441 AA 1442
Db 1364 AA 1365

```

RESULT 10
 ABL11658
 ID ABL11658 standard; cDNA; 3965 BP.
 XX
 AC ABL11658;
 XX
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 29456.
 XX
 XX Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.

```

XX XX
PN WO200171042-A2.
XX XX
PD 27-SEP-2001.
XX XX
PF 23-MAR-2001; 2001WO-US09231.
XX XX
PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX
PA (PEKE ) PE CORP NY.
XX XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX XX
DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB67555.
XX PT
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS
XX PS Claim 1; SEQ ID NO 29456; 21bp + Sequence Listing; English.
XX XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL1617-ABL30511), expressed DNA
CC sequences (ABL101840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 3965 BP; 1088 A; 946 C; 979 G; 952 T; 0 other;

```

Query Match 5.9%; Score 108.8; DB 23; Length 3965;
 Best Local Similarity 48.0%; Pred. No. 2.7e-20;
 Matches 462; Conservative 0; Mismatches 467; Indels 33; Gaps 4;

```

Qy 493 GAACAAATATGACGTAAGATTTCTGATTAAGCCGAGATATATCTTATGTTGTGCG 552
Db 1425 GAAGACTCCGAGGCGAGGACGAGTATCAAGCCAGACCACTTCTATGATGGGT 1484
Qy 553 CGAGCTGAACAGACGAGTGCATTTAGAGTGCATTTATATATGAAGAAGACTCT 612
Db 1485 CAGTTCAAGACGACGCGCTTCATGAGAGTGTGGTTTCAACGAGAGGAGAGGCT 1544
Qy 613 TTTTATGTAACACCATGATATCTCTGTGCAATATCTCTGAGTGTGAAATGCGTGAAT 672
Db 1545 CTCTACACCCACGACGATTTCTGTGCAAGCTTTCTCTGTGATCGATGATGATGAAT 1604
Qy 673 TTTGATCTAGCCAGATGATTTCTACTGAAATTAATCATTTGCTGTAGAAAACATGACCCCT 732
Db 1605 CAGGACGCGGCGAGCGAAAA---GGCGGCAACATGTGCGCATCGGCTGATGATCCG 1661
Qy 733 GTTATTGAAGTGTGGACCTTTGATATATGTGACTCTTTAGAGCCAGTCTTACACTCGGA 792
Db 1662 ATATACAGCTCTGGGATCTTAGACATACAGACGCTATGAGGCCCATTTAAGCTGGGT 1721
Qy 793 AGTAACTTTCAAAAGAAAAGAAAAGAAAAGAAAGAGTCTCTACAGACAGAGGCAT 852
Db 1722 TCCAAAGGACCGGAGACGAAACAAAGAA-----GCAATATGACAC 1763
Qy 853 ACCGATGCTGTCTTGAACCTTTATGGAATAGCTAATCAGAAATGTTTATGACAGTGA 912
Db 1764 AAGGAGCGCGTGTGATCTCTTGTGAAACCAACTTTAGACATTTTGGCCAGGGG 1823
Qy 913 TCAAGTGAACAACGTATTTCTGTGGATATGTCTTTGGGAAAACGAGACGTAGCTTC 972
Db 1824 TCCGTGACCAAACTGTGATTTCTGTGGACATGAGAGGCGGACGCTCATACCACTT 1883

```


QY 973 GCTGTACACAGACAGAGTCCAAACACTGACGTTTCATTCATTGAAGCAGACACTCTG 1032
 Db 1884 ACCGCTTTTGGCAAAACAGATTCACTGCGCTGGAATTCCTCCGCAAGAGCTCAAAAGCATT 1943
 QY 1033 ATTTCGCTCATATGATTAAGTCAGTGGCTTTGATGATGCTGCCAAGCTCAGATGAAGC 1092
 Db 1944 CTTACCGGCTGTGCCGATGATACGTGGCACTCTTCGATTGCCGACGCTGAGGGCTC 2003
 QY 1093 CA-----TCGAATGTGCGCATTCAGTGGCAGATGAGAGAGTGAATTCAC 1143
 Db 2004 AACTGTCACAGATTGAGTGAAGTTGACGCTGAAGAGGAGGAGTCTGTGGCATCC 2063
 QY 1144 TTTTACCTTGTCACTTTTGTGGCAGTACAGATGACGCTTTGATATATTTGGATGCA 1203
 Db 2064 ACACAGACCGACTACTTCATGCTGGGACCAACGATGGCACCCTTGATACGCCGACAAA 2123
 QY 1204 CGTTC---AGATAAGCCAAATTTTACATTAATGACACAAATGAAATCTGTGCTT 1260
 Db 2124 CGTTCCTGGAACAAGCTGTGTGCTCGTAAAGGCCCAACGAGAAATCTCCGGTGTG 2183
 QY 1261 GATCTTACAGTCAAAATCAAGGGCTGTCTCGTACCTGCTGACGTGACAAATACGTGAAG 1320
 Db 2184 TGTCTTCAACAAACAGAAACCTTAATCTGTAACCTCCACCTCCAGGAGGAGCACTTAAG 2243
 QY 1321 ATCTGGACATCTTAGAGATAGGCCAAAGTCTTAATCTTAGGAGACATGAATGGGA 1380
 Db 2244 GTGTGAACCTTGAATGACAGAGGCAAGCAAGCTTACGAGCAAGATTCACATGGGT 2303
 QY 1381 GTTCTCTTCTTCTTCTTCAATGTTGCCCTGATTTGCCATTATTTAATGCTTTGAGGTCA 1440
 Db 2304 CGCTTGACGTCAATGCGCCAGTGTCCGAGAGATCCCTACACCTGCGCTTGCGCGAGAG 2363
 QY 1441 AA 1442
 Db 2364 AA 2365
 RESULT 11
 ABL13260
 ID ABL13260 standard; cDNA; 4826 BP.
 AC ABL13260;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34262.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 DR P-PSDB; ABB69157.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 34262; 21pp + Sequence Listing; English.

XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16175-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB16173-AB172072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 4826 BP; 1258 A; 1202 C; 1196 G; 1170 T; 0 other;
 Query Match 5.9%; Score 108.8; DB 23; Length 4826;
 Best Local Similarity 48.0%; Pred. No. 3.1e-20;
 Matches 462; Conservative 0; Mismatches 467; Indels 33; Gaps 4;
 QY 493 GAACAAATGAACTGTAAGATTTCTTATTAAGCCCAAGTATATCTTATAGTTTGGC 552
 Db 369 GAAAGACTCCGAGCGCGAGAGAGAGTATCAAGCCGACAACTCATTTAGTGGT 428
 QY 553 CGAGCTGAACAGACGACAGTGAATTTAGAGTGCATGTTATATGAAGAAGACTCT 612
 Db 429 CACGTTGAAGACGAGCGCCCTCCATGAGGTGTGGATTTTCAACGAGAGAGAGGCT 488
 QY 613 TTTATGATACACATGATATATCTTGTCTGATATCTCTGAGTGTGAATGGCTGAAT 672
 Db 489 CTTCACACCAACAGACGACTTCTGCTCCTCAAGCTTCTCTGTCATCGATGATGAAT 548
 QY 673 TTTGATCTTGAACCCAGATGATTTCTACTGAAATTAATCACTTGTGTGAGAAACATGACCCCT 732
 Db 549 CACGACCGCGGCGAGCGAAAA--GGCGGCAACATGTGCCCATCGGCTCATGATGATCG 605
 QY 733 GTTATGAAGTGGGACCTTGATATAGTGAGACTCTTAGAGCCAGCTTCAACTCGGA 792
 Db 606 ATTAATCAAGCTGTGGATCTAGACATACAGAGCCCTTCAAGCCCATTTAAGTGGT 665
 QY 793 AGTAACTTTCAAAAAAGAAAAAGAAAAAGAAAGAGTTCTTACAGAGAGGCAAT 852
 Db 666 TCCAAAGGACGCGGAGAGCAAGAACAAAA-----GCAATGAGACAC 707
 QY 853 ACCGATGCTGTCTTGAACCTTTTCAATGATATAGCTATCAAGAAATGTTTAGCAAGTGA 912
 Db 708 AAGAGCCCGGTGCTGATCTCTTGTGAACCACTTTAGACACATTTGCGCAGGGGG 767
 QY 913 TCAGCTGACAACTGTAATCTGTGGGATATGTCCTTGGGAAAAACAGACGTAAGCTC 972
 Db 768 TCGGTGACCAACTGTGATTTCTGTGGACATGACAGAGGCGACCTCATACAACTAT 827
 QY 973 GCTGTACACAGACAGAGTCCAAACACTGCACTTTCATTCATTGAAGCAGACTCTG 1032
 Db 828 ACCGCTTTTGGCAAAACAGATTCACTGCTGGAATTCATCCGCAAGAGCTCAAAAGCAT 887
 QY 1033 ATTTCGCTCATATGATTAAGTCAGTGGCTTTGATGATGCTGCCAAGTCCAGATGAAGC 1092
 Db 888 CTTACCGGCTGTGCCGATGATACGTGCACTTTCGATTGCCGACGCTGAGGGGCTC 947
 QY 1093 CA-----TCGAATGTGCGCATTCAGTGGCAGATGAGAGAGAGTGAATTCAC 1143
 Db 948 AACTGTCACAGATTGAGTGAAGTTGACGCTGTAAGTGAAGAGTCTGTGGCATCC 1007
 QY 1144 TTTTACCTTGTCACTTTTGTGGCAGTACAGATGACGCTTTGATATATTTGGATGCA 1203
 Db 1008 ACACAGACCGACTACTTCATGTTGGGACCAAGATGGAACCTTGATACGCCGACAAA 1067
 QY 1204 CGTTC---AGATAAGCCAAATTTTACATTAATGACACAAATGATGAATCTGTGCTT 1260
 Db 1068 CGTTCCTGGAACAAGCTGTGTGCTCGTAAAGGCCCAACAGGAAATCTCCGGTGTG 1127
 QY 1261 GATCTTACGATCAAAATCAAGGGCTGTCTCGTACTGCTTACGTTCAATAATAGTGAAG 1320

Db 1128 TGCTTCAACAACGAAGCCTTAATCTGCTACCTCCACCTCCAGGAGGCCCTTAAG 1187
QY 1321 ATCTGGACATCTTAGAGATAGGCCAAGTCTAGTTCATTCTAGGACATGAAATGGA 1380
Db 1188 GTGTGAACTTTATGCGACAGAGCAAGACACTTACAGACGACGATTCACATGGGT 1247
QY 1381 GTTCTCTCTGTTCTTTCATGCTGCGCCCTGATTTGCCATTATTATGCTTGAAGTCAA 1440
Db 1248 CGCTTGACATGATGCGCCGACGATGCCCGAGATCCTACACCTCGCCTTCGCGGAGAG 1307
QY 1441 AA 1442
Db 1308 AA 1309

RESULT 12
AAC51036
ID AAC51036 standard; DNA; 1566 BP.
XX
AC AAC51036;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67037.
XX
KW Hybridization assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 22-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154739.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 16-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161350.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.7%; Score 106; DB 21; Length 1566;

Best Local Similarity 48.2%; Pred. No. 1e-19;
Matches 378; Conservative 0; Mismatches 385; Indels 21; Gaps 2;
QY 460 GATCAGATCCTTACGTTACTGGAAGATACAGAACATATGAACTGGAATTTCTTG 519
Db 460 GATCATTATCGAAAGATGTTGATGATGAAGATGATGAAGATATTTGATGACACGACA 519
QY 520 ATTAAAGCCAGTATATCTTATATGTTTGGCCGAGCTGAACAGACCGATGCAATTTA 579
Db 520 GTCAAGCCGACCGATTGGGATTTATTTGCTAAGAAATGAATGATGTTAGCATCTA 579
QY 580 GAGGTGATGTTTATTAATCAAGAAAGACTCTT-----TTATGTACACCATATGATA 633
Db 580 GAGGTATTTATTTATCGAGAAATGCTGCTGCTTCCAAACATGATGTTTATCATCATACATA 639
QY 634 CTCTTGTCTGATATCCTCTGAGTGGAAATGGCTGAATTTGATCTTACCCCATGATAT 693
Db 640 ATTATACAGAAATCCCATTTGTTACTGATGATGCTTATTTGCTTAAAGAGGGGAA 699
QY 694 TCTACTGGAATTCATTTGCTGTGAGAAACATGACCCCTGTTATTTGAAGTGGACCTT 753
Db 700 AAAGGAATTTTGTAGCTATTTGTTCAAGATACACCAATCGAATATGGATCTT 759
QY 754 GATATAGTGACCTCTTAAAGCCAGCTTTCACACTCGAAGTAACT----- 800
Db 760 GACGTTAGGAGCGAGTGTCTTACTTGTATACACTAGAGAAATAGAGATATAGT 819
QY 801 --TTCAAAAAGAAAGAAAAGAAAGAAAGAGATTCTCAGCAGAAAGGCATACGAT 858
Db 820 AGTAAGAAAAGAAAGAACAGAAACAGAAACCGAAATTTCAAGAAAGTACATATGAG 879
QY 859 GCTGTCTTGACCTTTATGATGAATACGTAATCAGAAATGTTTACGAAGTCATAGCT 918
Db 880 TCAGTACTTGTCTTGTCTTGTGAAACAGAGTTTCGAAACATCTGTAGTGTAGTGC 939
QY 919 GACAAACCTGTAATTTCTGTGGATATGCTTGGGAAACCAAGCAGCTAGCTCGCTGA 978
Db 940 GACAAAAGTTAAGTCTGTGGATGTGCTTACTGGAAGCTGTAAGATTTATATGAGACAT 999
QY 979 CACACAGCAAGTTCACAAACATGCACTGATTCATCTTGAAGCAGACTTGATTTCT 1038
Db 1000 CACAAAAGGAGTTCAAGCGGTTGCTTGAACCATTTATGCTCAGAAAGTCTTCAGT 1059
QY 1039 GCGTCAATGATTAAGTCAAGTGGCTTTGTATGACGTCCGAAGTCCAGATGAAGCATCGA 1098
Db 1060 GGGTCGTTTGAATCAACGTTGTGTATGAAGACGGAAGACAAACCTTCAACATCGGGTTTC 1119
QY 1099 ATGTGGCATTCAGTGGGCAGATGAGAGTGACTTGAATCACTTTTCACTTGTCTAT 1158
Db 1120 AAATGCTGTCAATGTCTGATGTGAAAGCTTAGCTGGATCTTCATTTGGAAACATCTCC 1179
QY 1159 TTCTTGGCCAGTACAGATGACGCTTTGTATATATTTGATGACGTTTCAGATAGCCA 1218
Db 1180 TTGTGTGATCTTTGAAGATGAAGACGTGAAGGTTTGTATATACGTCGACAGTCA 1239
QY 1219 ATTT 1222
Db 1240 GGT 1243
RESULT 13
AA21042
ID AA21042 standard; DNA; 553 BP.
XX AC AA21042;
XX 05-MAY-1999 (first entry)
XX Polynucleotide sequence from the genome of Treponema pallidum.
XX DE Treponema pallidum infection; syphilis; Borrelia infection; animal;
XX KW enzyme production; ds.
XX

OS Treponema pallidum.
XX WO9859034-A2.
XX 30-DEC-1998.
XX 23-JUN-1998; 98WO-US13041.
XX 24-JUN-1997; 97US-0050667.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Fraser CM;
XX WPI; 1999-081273/07.
XX New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
XX
XX Claim 1; Page 1034-1035; 1150pp; English.
XX
XX AAX0500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
XX Sequence 553 BP; 144 A; 105 C; 150 G; 153 T; 1 other;
SQ
XX
XX Query Match 5.6%; Score 104; DB 20; Length 553;
Best Local Similarity 80.3%; Pred. No. 2.1e-19;
Matches 122; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY 1238 ACAATGATGAATCTCTGCTTGTATCTTACGATCAATCAAGGGCTGTCTGTGACTG 1297
DB 34 ACACGTCTCTTCCCTTTAGGGCTGTGATCTCAGCAGTCAGATCAAGGGCTGTCTGTGACTG 93
QY 1298 CTTTCACTGACAAATACGTGAAGATCTGGGACATCTTAGAGATAGGCCAAGCTAGTTC 1357
DB 94 CATCAGCAGCAAAATACGTGAAGATCTGGGACATCTGGGAGACAGGCCGAGTCTGCTTC 153
QY 1358 ATTCTAGGACATGAATAATGGAGTCTCTTC 1389
DB 154 ATTCGCGGACATGAATAATGTAAGAAATCTCC 185
RESULT 14
AAS70865
ID AAS70865 standard; cDNA; 261 BP.
XX
XX AAS70865;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #6669.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA

XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABC06678.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1, SEQ ID No 6669; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 261 BP; 63 A; 73 C; 75 G; 50 T; 0 other;
SQ
XX
XX Query Match 4.0%; Score 74; DB 23; Length 261;
Best Local Similarity 100.0%; Pred. No. 6.2e-11;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 88 ATGACCGCAGCCGCCAGGTGACGTGCGTGGCTCGCTGCGCGCTGCGCAAGAG 147
DB 1 ATGAAACCGCAGCCGCCAGGTGACGTGCGTGGCTCGCTGCGCGCTGCGCAAGAG 60
QY 148 ACACGACGACAAAGT 161
DB 61 ACACGACGACAAAGT 74
RESULT 15
ABV17865
ID ABV17865 standard; cDNA; 211 BP.
XX
XX ABV17865;
AC
XX
XX 13-SEP-2002 (first entry)
DT
XX
XX Human prostate expression marker cDNA 17856.
DE
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
OS
XX WO200160860-A2.
PN
XX
XX 23-AUG-2001.
PD
XX
XX 20-FEB-2001; 2001WO-US05171.
PF
XX
XX 17-FEB-2000; 2000US-18319P.
PR 16-MAR-2000; 2000US-18962P.
XX
XX

PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 XX

PI Schlegel R, Endege WO, Monahan JE;
 XX

DR WPI; 2001-662795/76.
 XX

XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX

PS Claim 1; Page 2951; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;

CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX
 XX

SQ Sequence 211 BP; 74 A; 41 C; 40 G; 44 T; 12 other;

Query Match 3.8%; Score 69.6; DB 23; Length 211;

Best Local Similarity 84.8%; Pred. No. 1e-09;

Matches 78; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 805 AAAAAGAAAGAAAGAAAGAAAGAGTCTTCAGAGAAAGGCGATACGATGCTGTC 864

DB 112 AAAAATTAATTAAACAAATTTTACAGAGTCTTCAGAGAAAGGCGATACGATGCTGTC 171

OY 865 CTTGACCTTTCATGGAATAGCTAATCAGAAA 896

DB 172 CTTGACCTTTCATGGAATAGCTAATCAGAAA 203

Search completed: March 18, 2003, 22:32:43
 Job time : 896 secs

GenCore version 5.1.4_p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2003, 19:24:19 ; Search time 9898 Seconds
(without alignments)
3031.950 Million cell updates/sec

Title: US-09-315-355A-47
Perfect score: 1853
Sequence: 1 gatccctgagcgtgtgagcag.....aaagtaattcctaacaacat 1853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1148.6	62.0	2539	11 AK009972	AK009972 Mus muscu
2	966.6	52.2	1040	9 AL515984	AL515984 AL515984
3	953.4	51.5	995	9 AL561661	AL561661 AL561661
4	950.6	51.3	1092	9 AL582812	AL582812 AL582812
5	938	50.6	983	9 AL515985	AL515985 AL515985
6	932	50.3	1073	13 BMS47947	BMS47947 AGENCOURT

7	917	49.5	942	9 AL549213	AL549213 AL549213
8	906	48.9	906	14 BQ228676	BQ228676 AGENCOURT
9	880	47.5	902	9 AL576190	AL576190 AL576190
10	868	46.8	1007	13 BMS52364	BMS52364 AGENCOURT
11	865	46.7	894	14 BQ948509	BQ948509 AGENCOURT
12	847.4	45.7	869	14 BQ224661	BQ224661 AGENCOURT
13	842.4	45.5	953	14 BQ643631	BQ643631 AGENCOURT
14	825.4	44.5	941	14 BQ646683	BQ646683 AGENCOURT
15	823.2	44.4	965	14 BQ651718	BQ651718 AGENCOURT
16	821.8	44.3	1113	14 BQ064709	BQ064709 AGENCOURT
17	820.8	44.3	860	11 BQ032127	BQ032127 Homo sapi
18	818.8	44.2	866	9 AU125142	AU125142 AU125142
19	818.2	44.2	1002	13 BMS62222	BMS62222 AGENCOURT
20	813.4	43.9	1248	11 AK011891	AK011891 Mus muscu
21	813.2	43.9	980	14 BMS10786	BMS10786 AGENCOURT
22	805.8	43.5	922	14 BQ881882	BQ881882 AGENCOURT
23	802.8	43.3	809	9 AL552984	AL552984 AL552984
24	797.8	43.1	806	14 BQ425859	BQ425859 AGENCOURT
25	794	42.8	875	12 BMS75532	BMS75532 602598947
26	791.6	42.7	1022	13 BMS62256	BMS62256 AGENCOURT
27	786.6	42.5	1105	14 BQ276324	BQ276324 AGENCOURT
28	783.8	42.3	817	14 BMS98058	BMS98058 UI-CF-EN1
29	782.8	42.2	887	11 BQ010921	BQ010921 Homo sapi
30	780.8	42.1	1079	14 BMS913142	BMS913142 AGENCOURT
31	778.2	42.0	783	14 BMS59354	BMS59354 AGENCOURT
32	778.2	42.0	783	14 BMS09231	BMS09231 AGENCOURT
33	778.2	42.0	783	14 BMS13030	BMS13030 AGENCOURT
34	778.2	42.0	1043	14 BQ070859	BQ070859 AGENCOURT
35	776.2	41.9	782	13 BMS556169	BMS556169 AGENCOURT
36	773.8	41.8	916	14 BQ647936	BQ647936 AGENCOURT
37	773.2	41.7	789	9 AA639765	AA639765 nq84a04.s
38	766.2	41.3	796	13 BMA423756	BMA423756 AGENCOURT
39	751.8	40.6	919	14 BQ652250	BQ652250 AGENCOURT
40	750.8	40.5	772	13 BMA52846	BMA52846 AGENCOURT
41	750	40.5	783	9 AU125056	AU125056 AU125056
42	747.2	40.3	786	14 BMS12667	BMS12667 AGENCOURT
43	747	40.3	777	14 BMS78566	BMS78566 UI-CF-DU1
44	747	40.3	983	12 BE796924	BE796924 601587284
45	745.8	40.2	783	9 AI281923	AI281923 qm26a08.x

ALIGNMENTS

RESULT 1
AK009972
LOCUS 2539 bp mRNA linear HTC 19-JUN-2002
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310058a11:homolog to PERIODIC TRYPTOPHAN PROTEIN 1
HOMOLOGY (KERATINOCYTE PROTEIN IEF SSP 9502), full insert sequence.
ACCESSION AK009972
VERSION AK009972.1 GI:12845101
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain: C57BL/6J) adult male tongue cDNA to mRNA, clone lib:RIKEN full-length enriched mouse cDNA library
clone:2310058a11.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159


```

Db 427 TCTTAGGCTATGCGCATGATGACGAGATGCAATGTCACCTTGAAACACGGAACA 486
Qy 498 ATATGACGTAAGATTTCTGATTAAGCCAGATATATTAAGTTGTGGCCGAGC 557
Db 487 ATATGACATGAAGATTTCTGATTAAGCCAGATGACACCTCATCTGTGGCCGTC 546
Qy 558 TGAACAGACGAGTGCATTTAGAGTGATGATTTATTAACAAGAAAGACCTTTTAA 617
Db 547 TGAACAGAGCAGTGCATTTAGAGTGATGATTTATTAACAAGAAAGAGTCTTTCTA 606
Qy 618 TGTACACCATGATATCTCTTGTCTGATATCTCTGAGTGTGGAATGGCTGAATTTGA 677
Db 607 CGTCCACCATGATATCTCTGCTGATACCTCTGAGGTGGAATGGCTGAATTTGA 666
Qy 678 TCTTAGCCGAGATTTCTACTGAAATTTACATTTGCTGTGGAAGTGAACCTGTAT 737
Db 667 TCTTAGCCGAGATGATCTAAGTGAATTAACATTTGCTGTGGAAGTGAACCTGTAT 726
Qy 738 TGAAGTGTGGACCTTGATATAGTGAATGATCTTTAGACAGCTCTTCACTCGAAGTAA 797
Db 727 CGAGATATGGACCTGATATAGTGAATCTCTTGAACAGCTCTTCACTCGAAGTAA 786
Qy 798 ACTTTAAAAAAGAGAAAAAGAGAAAGAGTTCTTCAGCAGAGGGCATACCGA 857
Db 787 GCTTCCAAAAAGAGAAAGAGAAAGAGAGAGTTCTTCAGCAGAGAGCATACTGA 846
Qy 858 TGTGTCTCTGACCTTCAATGAAATAGCTATACAGAAAGTTTACAGTGCATAGC 917
Db 847 TGTGTCTCTGACCTTCAATGAAATAGCTATACAGAAAGTTTACAGTGCATAGC 906
Qy 918 TGAACAACACTGTAATCTGTGGAGATATGCTTGTGGGAAACAGCAGCTGCTGT 977
Db 907 AGACAGACACTGTAATCTGTGGAGATATGCTTGTGGGAAACAGCAGCTGCTGT 966
Qy 978 ACAACAGACAGAGTCCAAACACTGCAATTTCACTTTGAAGACAGACTGTATTC 1037
Db 967 ACATACAGACAGAGTCCAAACACTGCAATTTCACTTTGAAGACAGACTGTATTC 1026
Qy 1038 TGGCTCATATGAAATGCTGAGTCTTGTATGATGCTGCCGAAGTCCAGATGAAGCCATC 1097
Db 1027 TGGCTCATATGAAATGCTGAGTCTTGTATGATGCTGCCGAAGTCCAGATGAAGCCATC 1086
Qy 1098 AATGTGACGATGATGAGGAGATGAGAGAGTGTGGAATCACTTTTCACTGTCA 1157
Db 1087 CAGTGTGCGGATGATGAGGAGATGAGAGAGTGTGGAATCACTTTTCACTGTCA 1146
Qy 1158 TTTCTTGGCCAGTACAGATGACGCTTTGTATATTAATTTGATGACAGTTCAATAGACC 1217
Db 1147 TTTCTTGGCCAGTACAGATGATGAGCTTTGTATATTAATTTGATGACAGTTCAATAGACC 1206
Qy 1218 AATTTTAACTTAATGACACAAATGTAATCTCTGCTTGTATTTGACAGTCAAT 1277
Db 1207 AATTTTAACTTAATGACACAAATGTAATCTCTGCTTGTATTTGACAGTCAAT 1266
Qy 1278 CAAAGGCTGCTGTGATGCTTCACTGCACTGCAATATAGTGAAGTCTGGAGCATTTAGG 1337
Db 1267 CAAAGGCTGCTGTGATGCTTCACTGCACTGCAATATAGTGAAGTCTGGAGCATTTAGG 1326
Qy 1338 AGATAGGCCAAGTCTAGTCTTCTAGGACATGAAATGGAGTTCTCTTGTCTTC 1397
Db 1327 GATAGGCCAAGTCTAGTCTTCTAGGACATGAAATGGAGTTCTCTTGTCTTC 1386
Qy 1398 ATGTGCTGCTGATTTGATTTATGCTTGTGAGTCTCAAAAAGAGGCTTGGGT 1457
Db 1387 GTGCTGTCTGATTTGATTTATGCTTGTGAGTCTCAAAAAGAGGCTTGGGT 1446
Qy 1458 CTGGGATATAGACAGTCTTCAAGTAATGAAGATTTGGAAGAGGAGGCTGT 1517
Db 1447 TTGGGATATAGACAGTCTTCAAGTAATGAAGATTTGGAAGAGGAGGCTGT 1506
Qy 1518 TCTTGGAGTGAAGAAATCATCTATTAAGTGGCCCTTTGGCAGCAGAGTCA--GA 1574

```

```

Db 1507 TATAGGCGCACAAAGTTTATCTGTAGCGGCCCTGTGGAGCAGAGTCCACAGCA 1566
Qy 1575 TACCCATGAGTCTTAATGAAGATCA--TCTAATTTCTGCTTACTTAAGTGGAAAT 1633
Db 1567 GACACCAATGAGTCTCTGACAAAGATCAGTTTCTGCTGTACTTACTTACTCCG--TT 1624
Qy 1634 TTAATAAGTTGGCTTAAAAATGTTTC 1658
Db 1625 TCAAAAGACTGCGCTTAAAAATGTCC 1649

RESULT 2
AL515984/c 1040 bp mRNA linear EST 13-FEB-2001
LOCUS AL515984 LTI NF011 NBC1 Homo sapiens cDNA clone CS0DA001Y04 3
DEFINITION prime, mRNA sequence.
ACCESSION AL515984
VERSION AL515984.1 GI:12779477
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1040)
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..1040
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1b="CS0DA001Y04"
/clone_1b="LTI_NF011_NBC1"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library is not normalized, but is the control for
the normalized libraries. Library was constructed by Life
technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 298 a 230 c 206 g 303 t 3 others
ORIGIN

Query Match 52.2%; Score 966.6; DB 9; Length 1040;
Best local similarity 98.1%; Pred. No. 9, 8e-222;
Matches 998; Conservative 1; Mismatches 16; Indels 2; Gaps 2;

```

QY 1048 GATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCAGATGAAAGCCATCGAATGTGGCA 1107
 Db 777 GATAAGTCAGTGGCTTTGTATGACTGCCGAAGTCAGATGAAAGCCATCGAATGTGGCA 718
 QY 1108 TTCAATGCGGAGATGAGAGAGTGAATCTTTCACCTTTGCAATTTCTTGCC 1167
 Db 717 TTCAATGCGGAGATGAGAGAGTGAATCTTTCACCTTTGCAATTTCTTGCC 658
 QY 1168 AGTACAGATGAGCGGCTTGTATATTAATTTGATGACGCTTCAGATTAAGCAATTTTACA 1227
 Db 657 AGTACAGATGAGCGGCTTGTATATTAATTTGATGACGCTTCAGATTAAGCAATTTTACA 598
 QY 1228 CTTAATGACACAAATGATGAATCTCTGCTTTGATCTTAGACGTCGAATGAGGCTGT 1287
 Db 597 CTTAATGACACAAATGATGAATCTCTGCTTTGATCTTAGACGTCGAATGAGGCTGT 538
 QY 1288 CTGCTGACTGCTTACGCTGACCAATAGTGAAGATCTGGACATCTTAGAGATGAGCCA 1347
 Db 537 CTGCTGACTGCTTACGCTGACCAATAGTGAAGATCTGGACATCTTAGAGATGAGCCA 478
 QY 1348 AGTCTAGTCAATCTTAGGACATGAATAAGGAGTCTCTTGTCTTCATGTTGGCCT 1407
 Db 477 AGTCTAGTCAATCTTAGGACATGAATAAGGAGTCTCTTGTCTTCATGTTGGCCT 418
 QY 1408 GATTGCGATTTATTTATGCTTTGAGAGTCAAAAGAAGGCTTCGGGCTGAGATATA 1467
 Db 417 GATTGCGATTTATTTATGCTTTGAGAGTCAAAAGAAGGCTTCGGGCTGAGATATA 358
 QY 1468 AGCAGAGTCTCTTCAAGTAATGAAGCATTTGGAGACAGAGAGGCTTGTCTTGGAGT 1527
 Db 357 AGCAGAGTCTCTTCAAGTAATGAAGCATTTGGAGACAGAGAGGCTTGTCTTGGAGT 298
 QY 1528 GCAAGAAATTCATCTAATAGTGGCCCTTTGGACAGAGAGTCAATACCAATGAG 1587
 Db 297 GCAAGAAATTCATCTAATAGTGGCCCTTTGGACAGAGAGTCAATACCAATGAG 238
 QY 1588 TCTTAATGAATCATCTAATTTCTGCTTACCTTAACCTGGAATTTTAAAAAGTTGSCC 1647
 Db 237 TCTTAATGAATCATCTAATTTCTGCTTACCTTAACCTGGAATTTTAAAAAGTTGSCC 178
 QY 1648 TAAAAATGTCATGCGTGGGACGACCAATGAGTGAATGAAACAAATTCATTTCTG 1707
 Db 177 TAAAAATGTCATGCGTGGGACGACCAATGAGTGAATGAAACAAATTCATTTCTG 118
 QY 1708 ACTGACATTCCTTTCTGGAACCTGGGCTGGACCAAAATATCCGCTTTTGTGCTGCTC 1767
 Db 117 ACTGACATTCCTTTCTGGAACCTGGGCTGGACCAAAATATCCGCTTTTGTGCTGCTC 58
 QY 1768 TTCAAGATGATGTTTGAAGGCTCTTGTGCAATTTTAAAAAAGATTAATAAAAA 1824
 Db 57 TTCAAGATGATGTTTGAAGGCTCTTGTGCAATTTTAAAAAAGATTAATAAAAA 1
 RESULT 3
 AL561661 995 bp mRNA linear EST 16-FEB-2001
 LOCUS AL561661.LTI_NF1010_BC2 Homo sapiens cDNA clone CS0DL008YK21 5
 DEFINITION AL561661.LTI_NF1010_BC2 Homo sapiens cDNA clone CS0DL008YK21 5
 ACCESSION AL561661
 VERSION AL561661.1 GI:12909310
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 995)
 AUTHORS Li, W. B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Pull-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source
 location/Qualifiers
 1. .995
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DL008YK21"
 /clone_1ib="LTI_NF1010_BC2"
 /sex="male"
 /ribose_type="B cells from Burkitt lymphoma"
 /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com

BASE COUNT 291 a 203 c 263 g 237 t others
 ORIGIN
 Query Match 51.5%; Score 953.4; DB 9; Length 995;
 Best Local Similarity 99.4%; Pred. No. 1.5e-218;
 Matches 987; Conservative 1; Mismatches 2; Indels 3; Gaps 3;

QY 24 TGCGGTGCTGCTCCCTCCCTATGAGCGCTGGTCTAGCGTGACCGCCCTTGACTTGA 83
 Db 1 TGCGGTGCTGCTCCCTCCCTATGAGCGCTGGTCTAGCGTGACCGCCCTTGACTTGA 60
 QY 84 GACCATGAACCGGACCGCCGAGAGTGAAGTGGCTGGCTGGCTGGCTGGCTGGCTGG 143
 Db 61 GACCATGAACCGGACCGCCGAGAGTGAAGTGGCTGGCTGGCTGGCTGGCTGGCTGG 120
 QY 144 AGAAGACCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 203
 Db 121 AGAAGACCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 180
 QY 204 AAAGAGAAATTTGCAAGAGAGTGGTGGCAGTGAATGAAGAGAGAGAGAGAGAGAG 263
 Db 181 AAAGAGAAATTTGCAAGAGAGTGGTGGCAGTGAATGAAGAGAGAGAGAGAGAGAG 240
 QY 264 AGAAGATGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 323
 Db 241 AGAAGATGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 300
 QY 324 TGACCCAGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 383
 Db 301 TGACCCAGAGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 360
 QY 384 ATATGATGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 443
 Db 361 ATATGATGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 420
 QY 444 GGTCTACGGAGTGAATGATCAAGATCTTACCTTCTGAAAGATACAGAAATATGA 503
 Db 421 GGTCTACGGAGTGAATGATCAAGATCTTACCTTCTGAAAGATACAGAAATATGA 480
 QY 504 ACGTGAAGATTTCTGATTAAGCCAGTGAATCTTATAGTTTGTGGCGAGCTGAACA 563
 Db 481 ACGTGAAGATTTCTGATTAAGCCAGTGAATCTTATAGTTTGTGGCGAGCTGAACA 540
 QY 564 GGACCAAGTGAATTTAAGGAGTGAATTTAATCAAGAGAGAGAGAGAGAGAGAGAGAG 623
 Db 541 GGACCAAGTGAATTTAAGGAGTGAATTTAATCAAGAGAGAGAGAGAGAGAGAGAGAG 600
 QY 624 CCATGATATACCTTCTGCTGATATCTCTGAGTGAATGAGTGAATTTGATCTTGA 683
 Db 601 CCATGATATACCTTCTGCTGATATCTCTGAGTGAATGAGTGAATTTGATCTTGA 660
 QY 684 CCCGATGATTTCTAGTGAATTTATCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 743
 Db 661 CCCGATGATTTCTAGTGAATTTATCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 720

QY 744 GTGGGACCTTGATATAGTGACTCTTTAGACCAAGTCTTCACACTCGGAAGTAACTTTC 803
DB 721 GTGGGACCTTGATATAGTGACTCTTTAGACCAAGTCTTCACACTCGGAAGTAACTTTC 780
QY 804 AAAAAAGAGAAAAAGAGAAAAAGAGAGTCTCTCAGCAAGAGGCGATCCGATGCTGT 863
DB 781 AAAAAAGAGAAAAAGAGAAAAAGAGAGTCTCTCAGCAAGAGGCGATCCGATGCTGT 840
QY 864 CCTTGACCTTCAATGGAATAGCTAATCAGAAATGTTTGAAGATGATCAGCTGACAA 923
DB 841 CCTTGACCTTCAATGGAATAGCTAATCAGAAATGTTTGAAGATGATCAGCTGACAA 900
QY 924 CACTGTAATCTGTGGGATATGTCTCTGGGAAACAGACAGTACCTCGCTGACAC 983
DB 901 CACTGTAATCTGTGGGATATGTCTCTGGGAAACAGACAGTACCTCGCTGACAC 958
QY 984 AGACAGGTCGAAACAGTCACTGATTCATT 1016
DB 959 AGAC-ARGTCGAAACAGTCACTGATTCATT 990

RESULT 4
ALS82812/c ALS82812 1052 bp mRNA linear EST 16-FEB-2001
LOCUS ALS82812 LTI_NFL010 BC2 Homo sapiens cDNA clone CSDDL008YK21 3
DEFINITION prime, mRNA sequence.
ACCESSION ALS82812
VERSION ALS82812.1 GI:12951167
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (baes 1 to 1052)
Li, W.B., Gruber, C., Jessee, J. and Polyes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
Location/Qualifiers
1..1052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDDL008YK21"
/clone_1b="LTI_NFL010_BC2"
/sex="male"
/note="Vector: B cells from Burkitt lymphoma"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT)-primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
life Technologies. Contact : Feng Liang life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : Eliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 279 a 237 c 205 g 303 t 28 others
ORIGIN

Query Match 51.3%; Score 950.6; DB 9; Length 1052;
Best Local Similarity 95.7%; Pred. No. 6.9e-218;
Matches 1004; Conservative 23; Mismatches 16; Indels 6; Gaps 5;

QY 747 GGAACCTTGATATAGTGACTCTTTAGACCAAGTCTTCACACTCGGAAGTAACTTTC 806
DB 1047 GGAACCTTGATATAGTGACTCTTTAGACCAAGTCTTCACACTCGGAAGTAACTTTC 869
QY 807 AAAAAAGAGAAAAAGAGAAAAAGAGAGTCTCTCAGCAAGAGGCGATCCGATGCTGTCT 866

DB 988 AAAAAAGAGAAAAAGAGAAAAAGAGAGTCTTCAGACAAA-GGATATCCGATGCTGTCT 930
QY 867 TGACCTTTCATGGAATAGCTAATCAGAAATGTTTAGCAAGTGCATCAGTCAACAC 926
DB 929 TGACCTTTCATGGAATAGCTAATCAGAAATGTTTAGCAAGTGCATCAGTCAACAC 870
QY 927 TGTAAATCTGTGGGATATGTCTCTGGGAAACAGACAGTACGCTGCTGTACACAGA 986
DB 869 TGTAAATCTGTGGGATATGTCTCTGGGAAACAGACAGTACGCTGCTGTACACAGA 811
QY 987 CAAGTTCGAAACAGTCACTGATTCATCATTGAGACAGACCTCTGATTTTCGCTCAT 1046
DB 810 CAAGTTCGAAACAGTCACTGATTCATCATTGAGACAGACCTCTGATTTTCGCTCAT 751
QY 1047 TGATTAAGTCACTGCTTTGATGATGCTGCGAAGTCCAGATGAAAGCCATGATGCG 1106
DB 750 TGATTAAGTCACTGCTTTGATGATGCTGCGAAGTCCAGATGAAAGCCATGATGCG 651
QY 1107 ATTCACTGGGAGATAGAGAGATGACTTGAATCACTTTTCACTTTGATTTTGGC 1166
DB 690 ATTCACTGGGAGATAGAGAGATGACTTGAATCACTTTTCACTTTGATTTTGGC 631
QY 1167 CAGTACAGATGAGCGCTTTGATATATTTGATGACCGTTCAGATTAAGCAATTTTTC 1226
DB 630 CAGTACAGATGAGCGCTTTGATATATTTGATGACCGTTCAGATTAAGCAATTTTTC 571
QY 1227 ACTTAATGCACAATGATGAATCTCTGGCTTGAATCTTGAATCAATCAAGGCTG 1286
DB 570 ACTTAATGCACAATGATGAATCTCTGGCTTGAATCTTGAATCAATCAAGGCTG 511
QY 1287 TCTCGTACGCTCTTCACTGACGACAAATACGTAAGATCTGGGATCTTAGAGATAGCC 1346
DB 510 TCTCGTACGCTCTTCACTGACGACAAATACGTAAGATCTGGGATCTTAGAGATAGCC 451
QY 1347 AAGCTTAATGATCTTCAAGGACATGAATAAGGAGTCTTCTGTTCTTCAAGTGGCC 1406
DB 450 AAGCTTAATGATCTTCAAGGACATGAATAAGGAGTCTTCTGTTCTTCAAGTGGCC 391
QY 1407 TGATTTGCCATTAATTTATGCTTTGAGAGTCAAAAAGAGGCTTCGGGCTGGGATAT 1466
DB 390 TGATTTGCCATTAATTTATGCTTTGAGAGTCAAAAAGAGGCTTCGGGCTGGGATAT 331
QY 1467 AAGCAGAGTCTCTTCACTGATTAATGACATTTGAGACGAGAGGCTTTCTTGGAG 1526
DB 330 AAGCAGAGTCTCTTCACTGATTAATGACATTTGAGACGAGAGGCTTTCTTGGAG 271
QY 1527 TGCAGAAATTAATCTATTATAGTGGCCCTTTGGCAGAGAGGCTCAAGATACCCATGCA 1586
DB 270 TGCAGAAATTAATCTATTATAGTGGCCCTTTGGCAGAGAGGCTCAAGATACCCATGCA 211
QY 1587 GTCTTAATGAGATCAATCTAATTTCTGCTTACCTTAACCTGGGAATTTTAAAGTTGCG 1646
DB 210 GTCTTAATGAGATCAATCTAATTTCTGCTTACCTTAACCTGGGAATTTTAAAGTTGCG 151
QY 1647 CTAAATATGTTCCATGCGTGGCAGCAACATGACAGATGCTGAAACACATTCATTTCT 1706
DB 150 CTAAATATGTTCCATGCGTGGCAGCAACATGACAGATGCTGAAACACATTCATTTCT 91
QY 1707 GACTGA-CATTCTTTCTGCACTGCGGTGGCAGCAACAAATATCCGGTCTTTGTGCTTC 1765
DB 90 GACTGA-CATTCTTTCTGCACTGCGGTGGCAGCAACAAATATCCGGTCTTTGTGCTTC 31
QY 1766 TCTTCA--GATGATGTTTGTAAAGCTC 1792
DB 30 TCTTCA-CGATGATGTTTGTAAAGCTC 2

RESULT 5
ALS15985 ALS15985 983 bp mRNA linear EST 13-FEB-2001
LOCUS ALS15985 LTI_NFL011_NBC1 Homo sapiens cDNA clone CSDDA001YJ04 5
DEFINITION prime, mRNA sequence.
ACCESSION ALS15985

VERSION AL515985.1 GI:12779478
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 983)
 AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..983
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DA001J04"
 /clone_1db="LTI_NFL011_NBC1"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH108"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
 BASE COUNT 284 a 203 c 258 g 233 t 5 others
 ORIGIN
 Query Match 50.6%; Score 938; DB 9; Length 983;
 Best Local Similarity 99.4%; Pred. No. 7,4e-215;
 Matches 934; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 499 TATGACGTGAAGATTTCTTGATTTAAGCCAGATATATCTTATGTTGTGGCCGAGCT 558
 Db 481 TATGACGTGAAGATTTCTTGATTTAAGCCAGATATATCTTATGTTGTGGCCGAGCT 540
 QY 559 GAACAGAGACCGATGATTTAGAGGTGATGTTATATCAAGAGAGACTCTTTTAT 618
 Db 541 GAACAGAGACCGATGATTTAGAGGTGATGTTATATCAAGAGAGACTCTTTTAT 600
 QY 619 GTACACCATATATCTCTGTCTGATATCTCTGATGTTGAGTGAAGTGAATTTGAT 678
 Db 601 GTACACCATATATCTCTGTCTGATATCTCTGATGTTGAGTGAAGTGAATTTGAT 660
 QY 679 CTTAGCCCAATGATTTCTGATGAAATTAATGCTGTAGAAACATGACCCCTGATT 728
 Db 661 CTTAGCCCAATGATTTCTGATGAAATTAATGCTGTAGAAACATGACCCCTGATT 720
 QY 739 GAAGTGTGACCTTGTATATGATGACCTTTAGAGCAGCTTTCACCTCGAGATGAA 798
 Db 721 GAAGTGTGACCTTGTATATGATGACCTTTAGAGCAGCTTTCACCTCGAGATGAA 780
 QY 799 CTTTCAAAAAAGAAAAAGAAAAAGAAAGAGATTCTCAGCAGAAAGGCATACCGAT 858
 Db 781 CTTTCAAAAAAGAAAAAGAAAAAGAAAGAGATTCTCAGCAGAAAGGCATACCGAT 840
 QY 859 GCTGTCTTACCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 918
 Db 841 GCTGTCTTACCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 QY 919 GACACACTGATATTTCTGTGGATATGCTCTTGGGAGAAC 958
 Db 901 GACACACTGATATTTCTGTGGATATGCTCTTGGGAGAAC 940
 RESULT 6
 BMS47947 1073 bp mRNA linear EST 20-FEB-2002
 LOCUS BMS47947
 DEFINITION AGENCOURT_6531563 NIH_MGC_124 Homo sapiens
 5', mRNA sequence.
 ACCESSION BMS47947
 VERSION BMS47947.1 GI:18782142
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1073)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Straubeberg, Ph.D.
 Email: cga@bs-femail.nih.gov
 Tissue Procurement: Invitrogen
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LHAM12735 row: e column: 05
 High quality sequence start: 109
 High quality sequence stop: 739.
 Location/Qualifiers
 1..1073
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5732716"
 /clone_1db="NIH_MGC_124"
 /tissue_type="hippocampus"
 /lab_host="DH108"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site 2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average

Db	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809</
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------

FEATURES		High quality sequence stop: 677.
source	location/Qualifiers	
	1. 906	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:6061590"	
	/clone_1db="NIH MGC 72"	
	/issue_type="melanotic melanoma"	
	/lab_host="DH10B (phage-resistant)"	
	/note="Organ: skin; Vector: pCMV-Sport6; Site_1: NciI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."	
BASE COUNT	267 a 175 c 203 g 261 t	
ORIGIN		
Query Match	48.9%; Score 906; DB 14; Length 906;	
Best Local Similarity	100.0%; Pred. No. 3.6e-207;	
Matches	906; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	417 TCTTGGTGAATCTCTCTGGGCTTTACGGCTTACGGGAGTAATGATCAAGATCCTTACT	476
Db	1 TCTTGGTGAATCTCTCTGGGCTTTACGGCTTACGGGAGTAATGATCAAGATCCTTACT	60
Qy	477 TACTCTGGAAGATACAGAACATATGAACGTAGAATTCTTTGATTAAAGCCAGATATA	536
Db	61 TACTCTGGAAGATACAGAACATATGAACGTAGAATTCTTTGATTAAAGCCAGATATA	120
Qy	537 TCTTATAGTTTGTGGCCGAGCTGAACAGAACCATGTCAATTAGAGTGATGTTATTA	596
Db	121 TCTTATAGTTTGTGGCCGAGCTGAACAGAACCATGTCAATTAGAGTGATGTTATTA	180
Qy	597 TCAGAGAGAAGACTCTTTTATGTACACATGATATCTCTTGTGTGATATCCCTGAG	656
Db	181 TCAGAGAGAAGACTCTTTTATGTACACATGATATCTCTTGTGTGATATCCCTGAG	240
Qy	657 TGTGGAAATGGCGAATTTTGATTCCTAGCCAGATGATCTTCACTGSAATTAATGCTGT	716
Db	241 TGTGGAAATGGCGAATTTTGATTCCTAGCCAGATGATCTTCACTGSAATTAATGCTGT	300
Qy	717 AGAAGAACATGACCCCTGTATTGAAAGTGTGGACCTTGATATATGAGACTCTTTAGAGCC	776
Db	301 AGAAGAACATGACCCCTGTATTGAAAGTGTGGACCTTGATATATGAGACTCTTTAGAGCC	360
Qy	777 AGCTTCACTCGGAAGTAACTTTCAAAAAAGAAAGAAAGAAAGAAAGAAAGTTC	836
Db	361 AGCTTCACTCGGAAGTAACTTTCAAAAAAGAAAGAAAGAAAGAAAGAAAGTTC	420
Qy	837 CTCAGCAAGGGCATACCGATGCTGCTCTGACCTTTCACTGGAATTAAGCTATCAGAA	896
Db	421 CTCAGCAAGGGCATACCGATGCTGCTCTGACCTTTCACTGGAATTAAGCTATCAGAA	480
Qy	897 TGTTTTACAGTGCATGAGTGCACACACTGTAAATTTGTGTGGATATGTCTTGGGGAA	956
Db	481 TGTTTTACAGTGCATGAGTGCACACACTGTAAATTTGTGTGGATATGTCTTGGGGAA	540
Qy	957 ACCAGACAGTACCTGCTGTACACACAGACAAAGTCCAAACACTGCACTTTCATTCATT	1016
Db	541 ACCAGACAGTACCTGCTGTACACACAGACAAAGTCCAAACACTGCACTTTCATTCATT	600
Qy	1017 TGAACACAGACTGATTTTGGGCTCATATGATATAGTACAGTGGCTTTTATGACTGCGG	1076
Db	601 TGAACACAGACTGATTTTGGGCTCATATGATATAGTACAGTGGCTTTTATGACTGCGG	660
Qy	1077 AAGTCAGATGAAGCCATCGAATGTGGCGATTCACTGTGGCGAGATAGAGAGAGTGACTTG	1136
Db	661 AAGTCAGATGAAGCCATCGAATGTGGCGATTCACTGTGGCGAGATAGAGAGAGTGACTTG	720
Qy	1137 GAATCACTTTTCACTTGTCTATTCTTGGCCAGTACAGATGACGGCTTTGTATATAATT	1196
Db	721 GAATCACTTTTCACTTGTCTATTCTTGGCCAGTACAGATGACGGCTTTGTATATAATT	780
Qy	1197 GGATGACAGTTCAAGTAAGCCAAATTTTAACTTAATGACACATGATGAATCTCTG	1256

```

Db      781  GAGTGAACGTTGAGTAAGCAATTTTACCTTAATGACACAAATGATGAAATCTCTGG 840
QY      1257 TCTTGAATCTTGAAGCAATCAATCAAGGCTGTCTCGTACTGCTTCAAGTCAAAATACGT 1316
Db      841  TCTTGAATCTTGAAGCAATCAATCAAGGCTGTCTCGTACTGCTTCAAGTCAAAATACGT 900
QY      1317 GAAGAT 1322
Db      901  GAAGAT 906

RESULT 9
AL576190/c 902 bp mRNA linear EST 16-FEB-2001
LOCUS      AL576190 LTI_NPL006_PL2 Homo sapiens cDNA clone CSDDI072YE01 3
DEFINITION prime, mRNA sequence.
ACCESSION  AL576190
VERSION     AL576190.1 GI:12938088
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 902)
AUTHORS     Li, W.-B., Gruber, C., Jesse, J., and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 Evry cedex - France
             Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source      1..902
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="CSDDI072YE01"
             /clone_1b="LTI_NPL006_PL2"
             /tissue_type="Placenta"
             /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
             was primed with a NotI-Oligo(dT)-primer. Five prime end
             enriched, double-stranded cDNA was digested with Not I and
             cloned into the Not I and Eco RV sites of the pCMVSPORT 6
             vector. Library was normalized. Library was constructed by
             Life Technologies. Contact: Feng Liang Life Technologies,
             a division of Invitrogen 9800 Medical Center Drive
             Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
             Email : fliang@lifestech.com URL :
             http://fulllength.invitrogen.com"

BASE COUNT  267 a 205 c 180 g 246 t 4 others
ORIGIN
Query Match 47.5%; Score 880; DB 9; Length 902;
Best Local Similarity 99.2%; Pred. No. 6,5e-201;
Matches 891; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY      885  GCGAATCAAAAATGTTTGAAGAGTGCATGAGTGAACAACATGTAATTTGAGGATAT 944
Db      898  GTTAATCAAAAATGTTTGAAGAGTGCATGAGTGAACAACATGTAATTTGAGGATAT 839
QY      945  GTCTTGGGGAACACAGAGTACTGCTCGCTGACACACAGACAGAGTCCAAACACTGCA 1004
Db      838  GTCTTT -GGGAACACAGAGTACTGCTCGCTGACACACAGACAGAGTCCAAACACTGCA 780
QY      1005 GTTTCATCCATTTGAAGACACAGACTGATTTTGGCTCATATGATAGTCAAGTGGCTTT 1064
Db      779  GTTTCATCCATTTGAAGACACAGACTGATTTTGGCTCATATGATAGTCAAGTGGCTTT 720
QY      1065 GTATGACTGCCGAAGTCCAGATGAAGACATGATGAGTGGCGATTCAAGTGGGAGATAG 1124
Db      719  GTATGACTGCCGAAGTCCAGATGAAGACATGATGAGTGGCGATTCAAGTGGGAGATAG 660
QY      1125 GAGAGTGAATGGAATCACTTTCACTTGTCTTCTTGCCACAGTACAGATGACGCTT 1184

```

```

Db      659  GAGAGTGAATGGAATCACTTTTCACTTGTCTTGGCCAGTACAGATGACGCTT 600
QY      1185 TGTATATTAATTTGGATGACGCTTCAATTAAGCAATTTTAACTTAATGACACAAATG 1244
Db      599  TGTATATTAATTTGGATGACGCTTCAATTAAGCAATTTTAACTTAATGACACAAATG 540
QY      1245 TGAATCTCTGCTCTTATCTTATAGCAGTCAATCAAGGCTGTCTGTAAGTCTTCAAG 1304
Db      539  TGAATCTCTGCTCTTATCTTATAGCAGTCAATCAAGGCTGTCTGTAAGTCTTCAAG 480
QY      1305 TGAACAATACGTGAAGATCTGGGACATCTTGAAGATAGGCCAAGTCTATCTTCTAG 1364
Db      479  TGAACAATACGTGAAGATCTGGGACATCTTGAAGATAGGCCAAGTCTATCTTCTAG 420
QY      1365 GGCATGAAGAAATGGAGTCTCTTCTGTTCTTATAGTATGTCGCTGATTTGCAATTA 1424
Db      419  GGCATGAAGAAATGGAGTCTCTTCTGTTCTTATAGTATGTCGCTGATTTGCAATTA 360
QY      1425 TGCCCTTGGAGGTCAAAAAGAGGCTTGGGCTTGGATATTAAGACAGTCTTCAAGT 1484
Db      359  TGCCCTTGGAGGTCAAAAAGAGGCTTGGGCTTGGATATTAAGACAGTCTTCAAGT 300
QY      1485 AAATGAAGCATTTGGAAGACGAGAGAGGCTTGTCTTGGAGTGAAGAAATTCATCTAT 1544
Db      299  AAATGAAGCATTTGGAAGACGAGAGAGGCTTGTCTTGGAGTGAAGAAATTCATCTAT 240
QY      1545 TAGTGGCCCTTTTGGAGACGAGGCTCAATATACCCCATGAGTCTTAATGAAGATCAT 1604
Db      239  TAGTGGCCCTTTTGGAGACGAGGCTCAATATACCCCATGAGTCTTAATGAAGATCAT 180
QY      1605 TAATTCCTGCTTACTCTTAATCTGGGAATTTTAAAGTGGCTTAATAATGTTCCATGCG 1664
Db      179  TAATTCCTGCTTACTCTTAATCTGGGAATTTTAAAGTGGCTTAATAATGTTCCATGCG 120
QY      1665 TGGCAGCAACCATGACAGAGTACTGAAACAATTCATTTCTGACTGACATTCCTTTCTG 1724
Db      119  TGGCAGCAACCATGACAGAGTACTGAAACAATTCATTTCTGACTGACATTCCTTTCTG 60
QY      1725 CAATCGGCTGGGACCAACAATAATCCGCTTTCCTGCTTCTTGAATGATGCTT 1782
Db      59  CAATCGGCTGGGACCAACAATAATCCGCTTTCCTGCTTCTTGAATGATGCTT 2

RESULT 10
LOCUS      BM552364 1007 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT 6543589 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5549650
ACCESSION  BM552364
VERSION     BM552364.1 GI:18790164
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1007)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-remail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed By: The I.M.A.G.E. Consortium (ILNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/ILNL at:
             http://image.llnl.gov
             Plate: LLM12260 row: 1 column: 11
             High quality sequence stop: 699.
             Location/Qualifiers
             1..1007

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5549650"
/clone_lib="NIH MGC 71"
/tissue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-SORT6; Site_1: NCI; Site_2: Sall; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb."
Average insert size 2.1 kb.
BASE COUNT      287 a      212 c      272 g      234 t      2 others
ORIGIN

Query Match      46.8%; Score 868; DB 13; Length 1007;
Best Local Similarity 98.6%; Pred. No. 5.1e-198;
Matches 885; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY  6 CTGAGCGGTGGAGAGAGTGGGTGGTGGTCCCTCCATGACGCTGTTTCTAGCGTG 65
DB  1 CTGAGCGGTGGAGAGAGTGGGTGGTGGTCCCTCCATGACGCTGTTTCTAGCGTG 60
QY  66 ACAGCGCTTGTACTTGAAGACCATGAAACCGAGCCGCGAGGTGACGTGCGTGGCT 125
DB  61 ACAGCGCTTGTACTTGAAGACCATGAAACCGAGCCGCGAGGTGACGTGCGTGGCT 120
QY  126 CCGCTGCGCGCTGGCCAAAGACACACAGACAGGTAGAGTGAATAAGAAAGTAAA 185
DB  121 CCGCTGCGCGCTGGCCAAAGACACACAGACAGGTAGAGTGAATAAGAAAGTAAA 180
QY  186 ACGCTCATTTCTGAGGGAAGGAGAAATTGCAAGAAAGGTGGTGGCAGTGAAGA 245
DB  181 ACGCTCATTTCTGAGGGAAGGAGAAATTGCAAGAAAGGTGGTGGCAGTGAAGA 240
QY  246 GGAGACAGGAGTCCCTTGAAGAGATGCGATGAGAGTCAAGCAGCCAGGACGCGCAAG 305
DB  241 GGAGACAGGAGTCCCTTGAAGAGATGCGATGAGAGTCAAGCAGCCAGGACGCGCAAG 300
QY  306 AGAGCGCTTGAAGATGTTGATCCACAGAGATGACAGACCGTTTATGATGATGAGTGGC 365
DB  301 AGAGCGCTTGAAGATGTTGATCCACAGAGATGACAGACCGTTTATGATGATGAGTGGC 360
QY  366 TGAGTACAGTCTTGAATTAATGATGAGGAGGATGACCAAGTGTGAGACTCTTGGTGA 425
DB  361 TGAGTACAGTCTTGAATTAATGATGAGGAGGATGACCAAGTGTGAGACTCTTGGTGA 420
QY  426 ATCTCTCTGGGTCTTACGGTCTTACGGAGTAAATGATCAAGATCTTACTCTGAA 485
DB  421 ATCTCTCTGGGTCTTACGGTCTTACGGAGTAAATGATCAAGATCTTACTCTGAA 480
QY  486 AGATACAGAACATATGAAACGTGAAGATTTCTGATTAAGCCAGTGAATCTTATAGT 545
DB  481 AGATACAGAACATATGAAACGTGAAGATTTCTGATTAAGCCAGTGAATCTTATAGT 540
QY  546 TTGTCGCGGAGTGAACGAGACCAAGTGAATTAAGGAGTGAATTTAATCAAGAA 605
DB  541 TTGTCGCGGAGTGAACGAGACCAAGTGAATTAAGGAGTGAATTTAATCAAGAA 600
QY  606 AGACTCTTTTATGACACCATGATATATCTTGTCTGATATCTCTGAGTGGAAATG 665
DB  601 AGACTCTTTTATGACACCATGATATATCTTGTCTGATATCTCTGAGTGGAAATG 660
QY  666 GCTGAATTTTGTATGATCCAGCCAGATGATTTAATGAAATTAATGAGTGGTGA 725
DB  661 GCTGAATTTTGTATGATCCAGCCAGATGATTTAATGAAATTAATGAGTGGTGA 720
QY  726 GACCCCTGTATTTGAAGTGTGGACCTTGATATAGTGAATCTTTAAGGCGAGCTTAC 785
DB  721 GACCCCTGTATTTGAAGTGTGGACCTTGATATAGTGAATCTTTAAGGCGAGCTTAC 780
QY  786 ACTGGAAGTAACTTTCAAAAAGAAAGAAAGAAAGAAAGAGAGTCTCTACAGACA 845
DB  781 ACTGGAAGTAACTTTCAAAAAGAAAGAAAGAAAGAAAGAGAGTCTCTACAGACA 840
QY  846 AGGCGATACCGATGCTGTC- TTGACCTTTCAATGAAATAGCTAATCAGAAATGTTTT 902

```

```

DB  841 AAGCGATACCGATGCTCTCTTGTAGCTTTCAATGAAATAGCTAATCCAGAAATGTTT 898

RESULT 11
LOCUS   BQ948509
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 894)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2613 row: n column: 08
High quality sequence start: 14
High quality sequence stop: 670.
Location/Qualifiers
1. 894
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6427975"
/clone_lib="NIH MGC_101"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOT87; Site_1: EcoRI; Site_2:
XhoI; CDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT      264 a      167 c      210 g      252 t      1 others
ORIGIN

Query Match      46.7%; Score 865; DB 14; Length 894;
Best Local Similarity 98.7%; Pred. No. 2.6e-197;
Matches 882; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY  330 AGAGATGACAGACGCTTGATGATGATGAGCTGGTGAAGACACTTAAGTAATTA 389
DB  1 AGAGATGACAGACGCTTGATGATGATGAGCTGGTGAAGACACTTAAGTAATTA 60
QY  390 TGAGGAAGGTGACCCAGATGCTGAGACTCTTGGAATCTCTTGGGCTTACGGTCTA 449
DB  61 TGAGGAAGGTGACCCAGATGCTGAGACTCTTGGAATCTCTTGGGCTTACGGTCTA 120
QY  450 CCGGATTAATGATCAAGATCTTACCTTAAAGATACAGAACATATGAAAGTGA 509
DB  121 CCGGATTAATGATCAAGATCTTACCTTAAAGATACAGAACATATGAAAGTGA 180
QY  510 AGATTCTTGAATTAAGCCAGTGAATCTTATATGTTTGGCCGAGCTGAACAGACA 569
DB  181 AGATTCTTGAATTAAGCCAGTGAATCTTATATGTTTGGCCGAGCTGAACAGACA 240
QY  570 GTGCAATTAAGTGAATGTTTAAATCAAGAAAGAGACTTTTATATGACCACTGA 629

```


D	b	241	GTGCAATTAGAGGTGCACGTTTATTATATCAGAAAGAAAGACTCTTTTATTAGTACACATGA	300
O	y	630	TATACTCTTGCTCGCATCTCTCTGAGTGTGAATGGCTGAATTTTGAATCTTAGCCACGA	689
D	b	301	TATACTCTTGCTCGCATCTCTCTGAGTGTGAATGGCTGAATTTTGAATCTTAGCCACGA	360
O	y	690	TGAATCTACTGGAATTTACATGCTGCTAGGAAACATGACCCCTGTTATTGAAGTGGGA	749
D	b	361	TGATCTTACTGGAATTTACATGCTGCTAGGAAACATGACCCCTGTTATTGAAGTGGGA	420
O	y	750	CCCTGATATTAGTGAAGCTCTTTAGAGCCAGCTCTTCAACTCGAAGTAACTTTCAAAAA	809
D	b	421	CCTGATATTAGTGAAGCTCTTTAGAGCCAGCTCTTCAACTCGAAGTAACTTTCAAAAA	480
O	y	810	GAAAGAAAAAGAAAGAAAGAAAGAGTTCCCTCAGCAGAAAGGCGATACCAGTCTGTCTTGA	869
D	b	481	GAAAGAAAAAGAAAGAAAGAAAGAGTTCCCTCAGCAGAAAGGCGATACCAGTCTGTCTTGA	540
O	y	870	CCCTTCATGGAATTAAGCTATCAGAAATGTTTATAGCAATGATCAGTGCAGCAACTGT	929
D	b	541	CCCTTCATGGAATTAAGCTATCAGAAATGTTTATAGCAATGATCAGTGCAGCAACTGT	600
O	y	930	AATTCTGTGGATATGTCCTTGGGAAACACAGCAGTACGCTGCTGTAACACACAGACA	989
D	b	601	AATTCTGTGGATATGTCCTTGGGAAACACAGCAGTACGCTGCTGTAACACACAGACA	660
O	y	990	GGTCCAAACACTGCACTTTCATCCATTTTGAAGCAGACAGATCTGATTTCTGGCTCATATGA	1049
D	b	661	GGTCCAAACACTGCACTTTCATCCATTTTGAAGCAGACAGATCTGATTTCTGGCTCATATGA	720
O	y	1050	TAACTCAGTGGCTTGTATGATCTGCGGAAGTCCAGATGAAACCATCGAATGTGGGATT	1109
D	b	721	TAACTCAGTGGCTTGTATGATCTGCGCGAAGTCCAGATGAAACCATCGAATGTGGGATT	780
O	y	1110	CAGTGGGAGATAGAGAGAGTGAATCTTGGATCACTTTTCACTTGTCAATTTCTTGGCCAG	1169
D	b	781	CAGTGGGAGATAGAGAGAGTGAATCTTGGATCACTTTTCACTTGTCAATTTCTTGGCCAG	840
O	y	1170	TACAGATGA - CGGCTTGTATATATTTTGGATGACAGCTTCAATTAAGCAATTT 1222	
D	b	841	TACAGATGACCGCTTGTATATATTTTGGATGACAGCTTCAATTAAGCAATTT 894	

RESULT	12
LOCUS	BQ224661
DEFINITION	BQ224661 869 bp mRNA linear EST 02-MAY-2002
ACCESSION	AGNCOCURT_7569692 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6043941 5', mRNA sequence.
VERSION	BQ224661
KEYWORDS	BQ224661.1 GI:20406061
SOURCE	EST.
ORGANISM	human. Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 869)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/- National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Stransberg, Ph.D.
COMMENT	

FEATURES	Location/Qualifiers
source	1. 869
	/organism="Homo sapiens"

```

/cdb_xref="taxon:6036"
/cclone="IMAGE:6033941"
/cclone_id="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pCMV-SF0R6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dr' primed. Full-length insert size 2.5 kb. Library enriched for average-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      257 a      172 c      233 g      207 t
ORIGIN

```

[illegible]

Qy 876 ATGGAATAGCTTAATCAGAAATGTTTAG 904
 Db 841 ATGGGAATAGCTTAATCAGAAATGTTTAG 869
 RESULT 13
 B0643631 953 bp mRNA linear EST 15-JUL-2002
 LOCUS AGENCOURT 8342961 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6268799
 DEFINITION 5', mRNA sequence.
 ACCESSION B0643631
 VERSION B0643631.1 GI:21767803
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 953)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: ggaabs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLCM2431 row: e column: 24
 High quality sequence stop: 627.
 Location/Qualifiers
 1. .953
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6268799"
 /clone_id="NIH_MGC_100"
 /rname="hepatocellular carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pOT8; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 Library."
 BASE COUNT 272 a 187 c 224 g 270 t
 ORIGIN
 Query Match 45.5%; Score 842.4; DB 14; Length 953;
 Best Local Similarity 98.3%; Pred. No. 7.2e-192;
 Matches 873; Conservative 0; Mismatches 11; Indels 4; Gaps 2;
 Qy 955 AAACGAGAGCTAGGCTGCTGTACACACAGACAGCTCAAACTGACGTTTATCA 1014
 Db 1 AAACGAGAGCTAGGCTGCTGTACACACAGACAGCTCAAACTGACGTTTATCA 60
 Qy 1015 TTGGAAGACAGACTCTGATTTCTGGCTCATATGATTAAGTCAAGTGGCTTTGTATGACTGC 1074
 Db 61 TTGGAAGACAGACTCTGATTTCTGGCTCATATGATTAAGTCAAGTGGCTTTGTATGACTGC 120
 Qy 1075 CGAAGTCCAGATGAAAGCATGATGATGCGATTCAGTGGCGACATAGAGAGTGAAT 1134
 Db 121 CGAAGTCCAGATGAAAGCATGATGATGCGATTCAGTGGCGACATAGAGAGTGAAT 180
 Qy 1135 TGGATACCTTTTACCTTGTATTTCTGGCAGATGATGAGGCTTTTATATAT 1194
 Db 181 TGGATACCTTTTACCTTGTATTTCTGGCAGATGATGAGGCTTTTATATAT 240
 Qy 1195 TTGATGACAGTTCAGATTAAGCAATTTTACATTAATGACACAGATGATGAATCTCT 1254

Db 241 TTGATGACAGCTTCAGATTAAGCAATTTTACCTTAATGACACAGATGATGAATCTCT 300
 Qy 1255 GGTCTTGATCTTACAGTCAATCAAGGCTGTCTGTGACCTTACGTTCAAAATAC 1314
 Db 301 GGTCTTGATCTTACAGTCAATCAAGGCTGTCTGTGACCTTACGTTCAAAATAC 360
 Qy 1315 GTGAAGATCTGGACATCTTACAGATTAAGGCAAGCTCAATCTTACGTTCAAAATAC 1374
 Db 361 GTGAAGATCTGGACATCTTACAGATTAAGGCAAGCTCAATCTTACGTTCAAAATAC 420
 Qy 1375 ATGGAGTCTCTTCTGTCTTCTCATGTTGCTGATTTGACATTTATATGCTTTTGA 1434
 Db 421 ATGGAGTCTCTTCTGTCTTCTCATGTTGCTGATTTGACATTTATATGCTTTTGA 480
 Qy 1435 GGTCAAAAAGAGGCTTGGGCTGTGGATATAGACAGCTCTTTCAGTAAATGAAGCA 1494
 Db 481 GGTCAAAAAGAGGCTTGGGCTGTGGATATAGACAGCTCTTTCAGTAAATGAAGCA 540
 Qy 1495 TTGGAAGACAGAGAGGCTTGTCTTGGAGTGCAGAAATCATCTATTAAGGCGCT 1554
 Db 541 TTGGAAGACAGAGAGGCTTGTCTTGGAGTGCAGAAATCATCTATTAAGGCGCT 600
 Qy 1555 TTGGAAGACAGAGAGGCTTGTCTTGGAGTGCAGAAATCATCTATTAAGGCGCT 1614
 Db 601 TTGGAAGACAGAGAGGCTTGTCTTGGAGTGCAGAAATCATCTATTAAGGCGCT 660
 Qy 1615 CTACCTTAACCTGGAAATTTTAAAGTTGAGCTTAAATGTTCCATGCGTGGCAGAC 1674
 Db 661 CTACCTTAACCTGGAAATTTTAAAGTTGAGCTTAAATGTTCCATGCGTGGCAGAC 720
 Qy 1675 CATCAAGATGACCTGAACCAATTCATCTGACGACATTCCTTCTCACTGCGCT 1734
 Db 721 CATCAAGATGACCTGAACCAATTCATCTGACGACATTCCTTCTCACTGCGCT 780
 Qy 1735 GGCACCAAAATATCCGCTCTTGTGCTT-GCTCTTCAATGATGCTTTGTAAGGCTCT 1793
 Db 781 GGCACCAAAATATCCGCTCTTGTGCTT-GCTCTTCAATGATGCTTTGTAAGGCTCT 840
 Qy 1794 TGTTCATTTCTTAAAGATTAATTAAGATTTTAAAGTAA 1841
 Db 841 TGTTCATTTCTTAAAGATTAATTAAGATTTTAAAGTAA 885
 RESULT 14
 B0646683 941 bp mRNA linear EST 15-JUL-2002
 LOCUS AGENCOURT 8342479 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6267952
 DEFINITION 5', mRNA sequence.
 ACCESSION B0646683
 VERSION B0646683.1 GI:21770855
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 941)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: ggaabs-remail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLCM2431 row: b column: 17
 High quality sequence stop: 681.
 Location/Qualifiers
 1. .941
 FEATURES
 SOURCE

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6267952"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notice="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGGCGG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      280 a      189 c      248 g      223 t      1 others
ORIGIN
Query Match      44.5%; Score 825.4; DB 14; Length 941;
Best Local Similarity 99.0%; Pred. No. 8,9e-188;
Matches 851; Conservative 0; Mismatches 7; Indels 2; Gaps 2.

```

OY	102	CCAGGTGACGTGCGGTGGCTGGGTCTCGCTGCGGGGTGGCCAAAGACACCAAGCAAGGT	161
Dp	1	CCTGGTGAOGTGGCGTGGCTGGGTCCGTGTGGGGGTGGCCAAAGACACCAAGCAAGGT	60
OY	162	AGAGCTGAGTAAAGAGAAGTAAAAAGCCTCATCTGTCGAGGCCAAAGAGAAGAAATTGCAGA	221
Dp	61	AGAGCTGAGTAAAGAGAAGTAAAAAGCCTCATCTGTCGAGGCCAAAGAGAAGAAATTGCAGA	120
OY	222	AGAAAGTGTGTGGCAGTGTATGAAAGAGAGACAGGCAAGTCTTCAAAAGATGGCATGCGAG	281
Dp	121	AGAAAGTGTGTGGCAGTGTATGAAAGAGAGACAGGCAAGTCTTCAAAAGATGGCATGCGAG	180
OY	282	TGCACGCACCCAGGACGCCCCCAAGAGAGACCCCTGAGAGATGGTACCCAGAGATGCACAG	341
Dp	181	TGCACGCACCCAGGACGCCCCCAAGAGAGACCCCTGAGAGATGGTACCCAGAGATGCACAG	240
OY	342	GACCGTTGATGATGATGAGTGAAGTGGCTGGCTGAGTACGACTTGAATTAATATGATGAGAGAGTGA	401
Dp	241	GACCGTTGATGATGATGAGTGAAGTGGCTGGCTGAGTACGACTTGAATTAATATGATGAGAGAGTGA	300
OY	402	CCCGATGCTGAGACTCTTGGTGAATCTCTCTTGGGCTTACGGTCTACGGAGTAAATGA	461
Dp	301	CCCGATGCTGAGACTCTTGGTGAATCTCTCTTGGGCTTACGGTCTACGGAGTAAATGA	360
OY	462	TCAAGATCCTTAACGTACTCTGGAAGAGACAGACCAATATGAAGTGAAGTTCTTGAT	521
Dp	361	TCAAGATCCTTAACGTACTCTGGAAGAGACAGACCAATATGAAGTGAAGTTCTTGAT	420
OY	522	TAAGCCCAAGTAAATCTTATATGTTTGGCCGAGCTGAACAGAGACCAAGTGCATTTAGA	581
Dp	421	TAAGCCCAAGTAAATCTTATATGTTTGGCCGAGCTGAACAGAGACCAAGTGCATTTAGA	480
OY	582	GGTGCATGTTTATAATCAAGAAGAAGACTCTTTTATGTACACCATGATATACTCTTGTG	641
Dp	481	GGTGCATGTTTATAATCAAGAAGAAGACTCTTTTATGTACACCATGATATACTCTTGTG	540
OY	642	TGCATATCTCTGAAGTGGAAATGGCTGAATTTTGAATCTAAGCCCAATGATTTCTACTGG	701
Dp	541	TGCATATCTCTGAAGTGGAAATGGCTGAATTTTGAATCTAAGCCCAATGATTTCTACTGG	600
OY	702	AAATTCACATTCCTGAGGAAGAACATGACCCCTGTTATGAAATGTGGGACCTGATATAGT	761
Dp	601	AAATTCACATTCCTGAGGAAGAACATGACCCCTGTTATGAAATGTGGGACCTTGTATATAGT	660
OY	762	GGACTCTTTAAGACCAAGTCTTCAACTCGGAAGTAACTTTCAAAAAAGAGAAGAAAAAGAA	821
Dp	661	GGACTCTTTAAGACCAAGTCTTCAACTCGGAAGTAACTTTCAAAAAAGAGAAGAAAAAGAA	720
OY	822	AGGAAAGAA-GAGTTCCTCAGCAGAAAGGCAATCCGATGCTGTCTCTTGACCTTTCATGGA	880
Dp	721	AGGAAAGAAAGTGTCTCCACAGCAAGAGGCAATCCGATGCTGTCTCTTGACCTTTCATGGA	780

Oy	881	ATAAGCTATAGAAAGTTTATCAAGTCATCAGCTGACCAACACTGTATCTT	-GTGG	939
Db	781	ATAAGCTATACAGAAATCTTTTACCAAGTCATCANTGACAAACACTGTATCTT	CTGGGGG	840
Oy	940	GATATGTCCTTGGGGAACC	959	
Db	841	GATATGTCCTTGGGGAAC	860	

RESULT 15	
BQ651718	
LOCUS	BQ651718
DEFINITION	BQ651718 965 bp mRNA linear EST 15-JUL-2002 AGENCOURT 8303996 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6269343 5' , mRNA sequence.

ACCENSION	B0651718	GI:21775890
VERSION	B0651718.1	EST.
KEYWORDS	human.	
SOURCE	Homo sapiens	
ORGANISM		

REFERENCE
1 (bases 1 to 965)
AUTHORS
NIR-MGC <http://mgc.nci.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.

Tissue Procurement: CGAP (Stanford)
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LILCM2442 row: 1 column: 16
 High quality sequence stop: 645.

```

location/Qualifiers
1. 965
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6269343"
/clone_1kb="NIH_MGC_100"
/tissue type="hepatocellular carcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/node="Organ: liver; Vector: pOT8; Site:1: XhoI; Site:2:
EcoRI, cDNA made by oligo-dT priming, Directionally cloned
into EcoRI/XhoI sites using the following 5 adaptor:
GGGACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald L. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

```

BASE COUNT	286 a	202 c	257 g	219 t	1 others
ORIGIN					
Query Match		44.4%;	Score 823.2;	DB 14;	Length 965;
Best Local Similarity		98.2%;	Pred. No. 3e-187;		
Matches 842;		Conservative	0;	Mismatches 14;	Indels 1; Gaps
Qy	91	AACGCGACCCGCGAGCGTGGTGGCTGGGTCGCGCGCGCGGCGCAAGAAGACA	150		
Db	1	AACGCGACCCGCGAGCGTGGTGGCTGGGTCGCGCGCGCGGCGCAAGAAGACA	60		
Qy	151	CCAGCAAGGTGAGCTGAGTAAAGAAAGTAAACGCTCATTTGTTAGGCAAAAGAG	210		
Db	61	CCAGCAAGGTGAGCTGAGTAAAGAAAGTAAACGCTCATTTGTTAGGCAAAAGAG	120		
Qy	211	AAATTGCAAGAGAGGTGTGTGCGAGTGAAGAGAGACAGGCACTCCTTCAGAAAT	270		
Db	121	AAATTGCAAGAGAGAGGTGTGTGCGAGTGAAGAGAGAGACAGGCACTCCTTCAGAAAT	180		

```
OY 271 GGCATGAGAGTGACCGCACCGCCAGAGAGCCCTGAGAGATGTGACCCA 330
DB 181 GGCATGAGAGTGACCGCACCGCCAGAGAGAGCCCTGAGAGATGTGACCCA 240
OY 331 GAGGATGACAGAGCGCTTGATGATGAGTGGCTGAGTACGACTTAGATAAATATGAT 390
DB 241 GAGGATGACAGAGCGCTTGATGATGAGTGGCTGAGTACGACTTAGATAAATATGAT 300
OY 391 GAGGAGGTGACCGAGATGCTGAGACTCTTGGAATCTCTTGGGCTTACGGCTTAC 450
DB 301 GAGGAGGTGACCGAGATGCTGAGACTCTTGGAATCTCTTGGGCTTACGGCTTAC 360
OY 451 GGGAGTAATGATCAGATCCTTACCTTCTGAAAGATACAGAAACAATATGAACGTGAA 510
DB 361 GGGAGTAATGATCAGATCCTTACCTTCTGAAAGATACAGAAACAATATGAACGTGAA 420
OY 511 GATTCTTGATTAAGCCAGTGAATCTTATAGTTGTGGCCGAGCTGAAACAGACCAG 570
DB 421 GATTCTTGATTAAGCCAGTGAATCTTATAGTTGTGGCCGAGCTGAAACAGACCAG 480
OY 571 TGCAATTTAGAGGTCAGTGTATATCAAGAGAGACCTTTTATATGTACCATGAT 630
DB 481 TGCAATTTAGAGGTCAGTGTATATCAAGAGAGACCTTTTATATGTACCATGAT 540
OY 631 ATACTTGTCTGATATCCTCTGAGTGTGAAATGGCTGAATTTGATCTAGCCAGAT 690
DB 541 ATACTTGTCTGATATCCTCTGAGTGTGAAATGGCTGAATTTGATCTAGCCAGAT 600
OY 691 GATTCTACTGGAATTTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGGGAC 750
DB 601 GATTCTACTGGAATTTACATTGCTGTAGGAAACATGACCCCTGTTATTGAAGTGGGAC 660
OY 751 CTGATATAGTGAAGCTTTAGAGCCAGTCTTCACTCGGAAGTAACTTCAAAAAAG 810
DB 661 CTGATATAGTGAAGCTTTAGAGCCAGTCTTCACTCGGAAGTAACTTCAAAAAAG 720
OY 811 AAGAAAAAGAAAGAAAGAGATTCTCAGCAGAAAGGCAATCCGATGCTGCTTGAC 870
DB 721 AAGAAAAAGAAAGGGAAGAGATTCTCAGCAGAAAGGCAATCCGATGCTGCTTGAC 780
OY 871 CTTTCATGGAATTAAGCTAATCAGAAATTTTTAGCAAGT-GCATGAGCTGACACACTGT 929
DB 781 CTTTCATGGAATTAAGCTAATCAGAAATTTTTAGCAAGTGGCATCACTGACAACTGT 840
OY 930 AATTCTGTGGATATGT 946
DB 841 AATTTCTGGGGGATAT 857
```

Search completed: March 19, 2003, 18:16:54
Job time : 9961 secs